

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 11:14:06 ; Search time 15776 Seconds
(without alignments)
11475.903 Million cell updates/sec

Title: US-10-049-137-1

Perfect score: 4177

Sequence: 1 atgagtagttgttgagag.....caactggtaaatcaagattc 4177

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of Hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:**

- 1: gb_ba:**
- 2: gb_htg:**
- 3: gb_in:**
- 4: gb_om:**
- 5: gb_ov:**
- 6: gb_pat:**
- 7: gb_ph:**
- 8: gb_pl:**
- 9: gb_pr:**
- 10: gb_ro:**
- 11: gb_sts:**
- 12: gb_sy:**
- 13: gb_un:**
- 14: gb_vi:**
- 15: em_ba:**
- 16: em_fun:**
- 17: em_hum:**
- 18: em_in:**
- 19: em_mu:**
- 20: em_om:**
- 21: em_or:**
- 22: em_ov:**
- 23: em_pat:**
- 24: em_ph:**
- 25: em_pl:**
- 26: em_ro:**
- 27: em_sts:**
- 28: em_un:**
- 29: em_vi:**
- 30: em_htg_hum:**
- 31: em_htg_inv:**
- 32: em_htg_other:**
- 33: em_htg_mus:**
- 34: em_htg_pln:**
- 35: em_htg_rod:**
- 36: em_htg_mam:**
- 37: em_htg_vrt:**
- 38: em_sy:**
- 39: em_htgo_hum:**
- 40: em_htgo_mus:**
- 41: em_htgo_other:**

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4177	100.0	4334	8	AF185578	AF185578 Arabidops
2	4156	99.5	4180	8	AF185577	AF185577 Arabidops
3	850.8	20.4	3204	6	AX507831	AX507831 Sequence
4	471.6	11.3	6398	5	BC046866	BC046866 Xenopus 1
5	436.4	10.4	4343	10	BC060721	BC060721 Mus muscu
6	427.4	10.2	6438	10	BC058578	BC058578 Mus muscu
7	423	10.1	6434	9	AK128741	AK128741 Homo sapi
8	421.4	10.1	6327	6	A56817	A56817 Sequence 1
9	421.4	10.1	6328	6	AR224019	AR224019 Sequence
10	421.4	10.1	6328	6	AR269336	AR269336 Sequence
11	421.4	10.1	6328	6	AR367436	AR367436 Sequence
12	421.4	10.1	6417	6	AX332238	AX332238 Sequence
13	421.4	10.1	6417	6	HSM12218	HSM12218 Sequence m
14	421.4	10.1	6475	9	AR338834	AR338834 Sequence
15	420.8	10.1	6331	6	AX924075	AX924075 Sequence
16	420.8	10.1	6331	6	AX924076	AX924076 Sequence
17	420.8	10.1	6331	9	AF066515	AF066515 Homo sapi
18	420.6	10.1	5689	3	AF308445	AF308445 Caenorhab
19	420.6	10.1	6771	6	AX924049	AX924049 Sequence
20	420.6	10.1	6771	9	HSU91543	U91543 Homo sapien
21	400.2	9.6	6584	9	BC038596	BC038596 Homo sapi
22	387	9.3	3716	6	AX746951	AX746951 Sequence
23	387	9.3	3716	9	AK091386	AK091386 Homo sapi
24	387	9.3	4563	9	HSM801006	AL117491 Homo sapi
25	387	9.3	9646	9	AF425231	AF425231 Homo sapi
26	385.4	9.2	4386	9	HSM802121	AL137693 Homo sapi
27	385.4	9.2	4432	9	HSM800877	AL110281 Homo sapi
28	374	9.0	18029	3	CEH06001	292970 Caenorhabdi
29	363.8	8.7	5603	3	AF308444	AF308444 Caenorhab
30	356.6	8.5	4966	9	AK127046	AK127046 Homo sapi
31	347.4	8.3	42467	9	U55373	U55373 Caenorhabdi
32	324.8	7.8	156251	2	AC107325	AC107325 Drosophi
33	324.8	7.8	183948	3	AC009376	AC009376 Drosophi
34	324.8	7.8	194308	3	AC009377	AC009377 Drosophi
35	324.8	7.8	257692	3	AE003517	AE003517 Drosophi
36	322.8	7.7	157766	2	AC146651	AC146651 Medicago
37	321.6	7.7	4912	3	AF007780	AF007780 Drosophi
38	321.2	7.7	89380	8	AP004259	AP004259 Oryza sat
39	318.2	7.6	81729	8	AB016874	AB016874 Arabidops
40	318.2	7.6	82359	8	ATAC002342	AC002342 Arabidops
41	317.2	7.6	92783	3	AC084459	AC084459 Caenorhab
42	313.2	7.5	6475	3	AF119716	AF119716 Drosophi
43	312.8	7.5	3079	3	AY071503	AY071503 Drosophi
44	311.6	7.5	150884	2	AC009382	AC009382 Drosophi
45	311.6	7.5	167816	3	AC009381	AC009381 Drosophi

ALIGNMENTS

RESULT 1

AF185578

LOCUS AF185578 4334 bp mRNA linear PLN 10-NOV-1999

DEFINITION Arabidopsis thaliana GYMNOSE/PICKLE mRNA, complete cds.

ACCESSION AF185578

VERSION AF185578.1 GI:6318929

KEYWORDS Arabidopsis thaliana (thale cress)

SOURCE Arabidopsis thaliana

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE 1 (bases 1 to 4334)

Ashed, Y., Baum, S.F. and Bowman, J.L.

AUTHORS Distinct mechanisms promote polarity establishment in carpels of

TITLE

JOURNAL	Arabidopsis thaliana	Db	305	GATAAGATATTGGATTGTGAAATGCGTCTCAAAAAATCTAGTGAACAAGGTTCTCCCGAT	364
REFERENCE	Cell (1999) In press	Qy	361	GCAGAACCGAAGCAATTTTGTGAAACAGTATCTCGTGAAGTGGAGGATATATCATAC	420
AUTHORS	2 (bases 1 to 4334)	Db	365	GCAGAACCGAAGCAATTTTGTGAAACAGTATCTCGTGAAGTGGAGGATATATCATAC	424
TITLE	Eshed,Y.	Qy	421	CTTCACTGCTCTTGGGTGCTGAGAGGAGTTCAGAAAGCTTTATAAGTCAAAATCATCGT	480
JOURNAL	Submitted (12-SEP-1999) Plant Biology, UC Davis, One Shields Ave., Davis, CA 95616, USA	Db	425	CTTCACTGCTCTTGGGTGCTGAGAGGAGTTCAGAAAGCTTTATAAGTCAAAATCATCGT	484
FEATURES	Location/Qualifiers	Qy	481	TTAAAAACAGAGTGAACAAATTTTCCCGTCAAAATGGAGTCAATTAATAACAGCGAAGAT	540
source	1..4334	Db	485	TTAAAAACAGAGTGAACAAATTTTCCCGTCAAAATGGAGTCAATTAATAACAGCGAAGAT	544
	/organism="Arabidopsis thaliana"	Qy	541	GATTTTGTTCCTACATGCTCTGAGTGGAGCCACCTGTTGATCGGATCTTTCCTCAGAGAG	600
	/mol_type="mRNA"	Db	545	GATTTTGTTCCTACATGCTCTGAGTGGAGCCACCTGTTGATCGGATCTTTCCTCAGAGAG	604
	/product="GYMNOS/PICKLE"	Qy	601	GAAGATGGGAGCTGGAATATCTTGTCAAAATATAAAGAGCTATCTTATGATGAATGTAT	660
	/protein_id="AAF07084.1"	Db	605	GAAGATGGGAGCTGGAATATCTTGTCAAAATATAAAGAGCTATCTTATGATGAATGTAT	664
	/db_xref="GI:6318930"	Qy	661	TGGGAGTCAGAAATCAGACATCTCAACCTTCAGAAATGAATTCAAAGGTTCAAGGATGTA	720
	/translation="MSSLVERLIRSDRKPVNLDSDDDDDFVPPKDDTPEOVEAIVR	Db	665	TGGGAGTCAGAAATCAGACATCTCAACCTTCAGAAATGAATTCAAAGGTTCAAGGATGTA	724
	TDKENACAGGSESTNLVSCNTFYAFHAKCLVPPPLKDSAVENWRCPECVSPANEIDK	Qy	721	AATTCTAGAACTCGCAGAAATGAAGATGTTGACCAATAAAGAAATCCAGAGACTTTCAA	780
	ILDCEMPTKSSGSDAEPKPIFVKQILVKKWGLSYLHCSWVRCPEKOKAYKSNHR	Db	725	AATTCTAGAACTCGCAGAAATGAAGATGTTGACCAATAAAGAAATCCAGAGACTTTCAA	784
	LKTRVNNFHRQESFNSSDDFAIREPTWTDRIACREEDGELEYLVKXELSYDE	Qy	781	CAGTTTGTATCATACTCTGAAATTCCTCAAGGCTTGTATACATCCATACCAAGCTTGAGGA	840
	CYWESEDI STFQNEIQRFDKVNSTRSDVDHKNRPDQFDPHTPEFKGLLHPY	Db	785	CAGTTTGTATCATACTCTGAAATTCCTCAAGGCTTGTATACATCCATACCAAGCTTGAGGA	844
	QLEGLNFRFSWKOTYHILADEMLGKTIQSIALLASLFEENLI PHLVIAPLSTLRN	Qy	841	CTTAAATTTTGGCGTCTCGTGGTCAAAAACAGACGATGTAATCTCTGCTGATGAAATG	900
	WEREFATWQPMVMVYFGTAQAPAVIREHEFYLSKDOKI KKKSKQI SSESQKRI	Db	845	CTTAAATTTTGGCGTCTCGTGGTCAAAAACAGACGATGTAATCTCTGCTGATGAAATG	904
	KFDVLLTSYEMINLDSAVLXPKIWECHMIVDEGHLKNKDSKLSFSLTQHSNRRILLT	Qy	901	GGACTAGGCAAGCAATTCAAAGCAATTCGCTTTTACCTTTTACCTTTTGGAGGAGAACCTC	960
	GTPVNDLDELFLMHLDAGKFGSLEEFQDFDINOEQI SRLHOMLAPHLLRRVK	Db	905	GGACTAGGCAAGCAATTCAAAGCAATTCGCTTTTACCTTTTACCTTTTGGAGGAGAACCTC	964
	KDVNMPPKKEILIRLDLSLQKEYKAI FTNRYQVLTGKGAQI SLNNIMBELRVK	Qy	961	ATTCCGCTATTGGTAAATGCTCTCTATCGACTCTCGTAACTGGGAGAGAGATTGGCC	1020
	DLLEDYCTHKWQYERIDGKVGGAEROIRIDRFNKNKNKFCFLSLTRAGGLGNLAT	Db	965	ATTCCGCTATTGGTAAATGCTCTCTATCGACTCTCGTAACTGGGAGAGAGATTGGCC	1024
	ADTVLIYDSDWNPHADLOMARAHRLGOTWKMVLYRLINRGTLIEERMMOLTKKMVL	Qy	1021	ACATGGGCCCCACAGATGAACGTTGTTATGTTTGGCACTCGGCAAGCTTCAGAGAGTT	1080
	HLVGLKLTQINQSELDLIIRYSGKELFASDEDEAGKSKIHYYDDDAIDKLDRDLY	Db	1025	ACATGGGCCCCACAGATGAACGTTGTTATGTTTGGCACTCGGCAAGCTTCAGAGAGTT	1084
	EABEVSVDDEENGLKAFKVPFVYIDENEAALAEQRAVAESKSSAGNSDRASYE	Qy	1081	ATCAGAGAACATCAGATTCTTACTTTATCGAAGATCAAAAAGATCAAGAAAAAGAAATCT	1140
	ELLKDFELQAEALNALGRKRSKQLVSI BEDDLAGLEDVSDGSEYAEASTDGE	Db	1085	ATCAGAGAACATCAGATTCTTACTTTATCGAAGATCAAAAAGATCAAGAAAAAGAAATCT	1144
	AAGCGVOTGRPRYKRDNLPTLMGEGRGRFVLGPNQSORAI PVQTLRYGAGN	Qy	1141	GGACAAATAGTACGGAAGCAAGCAAAAAGAAATCAAGTTTGATGTCTCTCACATCG	1200
	FWKEFPVRLKQTFEINEYGLFLKHAIEEDENSPTFDGVPKELRI EDVLVRI	Db	1145	GGACAAATAGTACGGAAGCAAGCAAAAAGAAATCAAGTTTGATGTCTCTCACATCG	1204
	ALLVIOEKVKEVEDHPKVPFSPRIILERPGLRSKGIWKEHDKIMIRAVLKHGVR	Qy	1201	TATGAGATGATCAACCTAGATTTCAGACGTTCTAAAAACCAATTAAGTGGGAGTGCATGATT	1260
	WQAIVDDELGIQELICKELNFPHISLAAEQAGLQONGSGSGNFGAQNQPGSVI	Db	1205	TATGAGATGATCAACCTAGATTTCAGACGTTCTAAAAACCAATTAAGTGGGAGTGCATGATT	1264
	TGNNASDAQVQSMFYRYDMQRRLVFEVKRVLLLEKAWNYEAYEEYXGLGSSSI	Qy	1261	GTGTGAGAGTCAACCTAGATTTCAGACGTTCTAAAAACCAATTAAGTGGGAGTGCATGATT	1320
	PTEPEAPKADTVGSFIEVDMDLGLPTDPTITSEEIMGAADVNDQARVEIAQH	Db	1265	GTGTGAGAGTCAACCTAGATTTCAGACGTTCTAAAAACCAATTAAGTGGGAGTGCATGATT	1324
	YNQCKLLDNARESQVAYVNPPTKVNESFRALKS INGNINILISITSDQSKHE	Qy	1321	TATTCAAGTAAACCCCGTATTTCTTCTGACAGGAACCACTTCAGAACCAACTTGGATGAA	1380
	DDTKPDLNNVEMKDTABETKPLRGVVLDLVVEEENIAEASGSDVDMKEEABEEKP	Db	1325	TATTCAAGTAAACCCCGTATTTCTTCTGACAGGAACCACTTCAGAACCAACTTGGATGAA	1384
	KNMVVD"	Qy	1381	CTTTTCATGCTCATGCAATTTTCTTGATCGCGGGAAGTTTGGAGGAGTTCCAG	1440
		Db	1385	CTTTTCATGCTCATGCAATTTTCTTGATCGCGGGAAGTTTGGAGGAGTTCCAG	1444
ORIGIN					
Query Match	100.0%; Score 4177; DB 8; Length 4334;				
Best Local Similarity	100.0%; Pred. No. 0;				
Matches 4177; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
Qy	1	ATGAGTAGTTGCTGGAGAGCTTCGCATACGATCTGATAGGAACCAAGTTTATAACCTA	60		
Db	5	ATGAGTAGTTGCTGGAGAGCTTCGCATACGATCTGATAGGAACCAAGTTTATAACCTA	64		
Qy	61	GATGATTTCTGATGATGACGATCTCGTCTCTAAAAAGATCGAACCTTTTGAGCAAGTCGAG	120		
Db	65	GATGATTTCTGATGATGACGATCTCGTCTCTAAAAAGATCGAACCTTTTGAGCAAGTCGAG	124		
Qy	121	GCTATTGTCAGAACTGATGCGAAAGAAAATGATGTCAGGCTTGTGGGAAAGTACTAAT	180		
Db	125	GCTATTGTCAGAACTGATGCGAAAGAAAATGATGTCAGGCTTGTGGGAAAGTACTAAT	184		
Qy	181	CTTGTAGCTGCATACATGCACTATGCGTTCATGCTAAATGCTTAGTTCACCTCTT	240		
Db	185	CTTGTAGCTGCATACATGCACTATGCGTTCATGCTAAATGCTTAGTTCACCTCTT	244		
Qy	241	AAAGATGCTTCGCTGGAAAAATGGAGATGCGCTCAATGTGTAGTCTCTTAAACGAGATA	300		
Db	245	AAAGATGCTTCGCTGGAAAAATGGAGATGCGCTCAATGTGTAGTCTCTTAAACGAGATA	304		
Qy	301	GATAAGATATTGGAAATGTGAAATCGCTCTCAAAAAATCTAGTGAACAAGGTTCTCCGAT	360		

Qy	1441	GAGGAGTTCAAAGATATTAATCAAGAGGAGCAGATCTCAAGGTTGCAAAAAATGTTGGCT	1500
Db	1445	GAGGAGTTCAAAGATATTAATCAAGAGGAGCAGATCTCAAGGTTGCAAAAAATGTTGGCT	1504
Qy	1501	CCACATTTGCTCAGAAGGGTAAAGAGAGCGTAATGAAGACATGCCCCCAAAAGGAG	1560
Db	1505	CCACATTTGCTCAGAAGGGTAAAGAGAGCGTAATGAAGACATGCCCCCAAAAGGAG	1564
Qy	1561	CTCATTTTGGGTGTGATCTGAGCAGTCTGCAGAAAGAAATATTACAAAGCTATTTTACC	1620
Db	1565	CTCATTTTGGGTGTGATCTGAGCAGTCTGCAGAAAGAAATATTACAAAGCTATTTTACC	1624
Qy	1621	CGTAATTATCAAGTATTGACAAAAAGGGAGGTGCTCAAAATTTCCCTTAATAACATTTATG	1680
Db	1625	CGTAATTATCAAGTATTGACAAAAAGGGAGGTGCTCAAAATTTCCCTTAATAACATTTATG	1684
Qy	1681	ATCGAATTACGAAAGTATGCTGCCATCTTATATGCTAGAGGGTGTGAGCCAGTTATT	1740
Db	1685	ATCGAATTACGAAAGTATGCTGCCATCTTATATGCTAGAGGGTGTGAGCCAGTTATT	1744
Qy	1741	CAGCAGCAAAATGAAGCTTTCAAACTTTTGGAGTCTTTGGAAAGCTGCAACTTCTA	1800
Db	1745	CAGCAGCAAAATGAAGCTTTCAAACTTTTGGAGTCTTTGGAAAGCTGCAACTTCTA	1804
Qy	1801	GATAAAATGATGTTCAAACTGAAAGAGCAAGGACACAGAGTCTCTAATATACACAGTTT	1860
Db	1805	GATAAAATGATGTTCAAACTGAAAGAGCAAGGACACAGAGTCTCTAATATACACAGTTT	1864
Qy	1861	CAGCATATGCTGGACTTACTTTGAAGACTACTGTACCCATGAAGAAATGGCAGTACGAGCA	1920
Db	1865	CAGCATATGCTGGACTTACTTTGAAGACTACTGTACCCATGAAGAAATGGCAGTACGAGCA	1924
Qy	1921	ATTGATGAAAGGTTGGCGAGCTGAGCGCAAAATACGCATAGATCGGTTCAATGCGCAA	1980
Db	1925	ATTGATGAAAGGTTGGCGAGCTGAGCGCAAAATACGCATAGATCGGTTCAATGCGCAA	1984
Qy	1981	AATTCATAAGTTTGTGTTTTTGTCTCCACAAGAGCTGGTGGCTTTAGGAATAAATCTT	2040
Db	1985	AATTCATAAGTTTGTGTTTTTGTCTCCACAAGAGCTGGTGGCTTTAGGAATAAATCTT	2044
Qy	2041	GCAACGCTGATACAGTAATTAATGACAGTCACTGGAATCTCATGCTGATCTCAA	2100
Db	2045	GCAACGCTGATACAGTAATTAATGACAGTCACTGGAATCTCATGCTGATCTCAA	2104
Qy	2101	GCAATGGCTAGAGCTCATCGACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCAT	2160
Db	2105	GCAATGGCTAGAGCTCATCGACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCAT	2164
Qy	2161	AACCGAGGCCAATTTGAAGAAAGATGATGCAATTTGACTAAAAAGAAATGGTTCTAGAG	2220
Db	2165	AACCGAGGCCAATTTGAAGAAAGATGATGCAATTTGACTAAAAAGAAATGGTTCTAGAG	2224
Qy	2221	CATCTTTGCTTGGGAACTCAAAACACAAACATTAATCAGNAGGTTAGATGACATC	2280
Db	2225	CATCTTTGCTTGGGAACTCAAAACACAAACATTAATCAGNAGGTTAGATGACATC	2284
Qy	2281	ATCAGGTATGGATCAAAAGGAGCTTTTGTAGTGAAGATGATGAAGCAGGAAAGTCTGGA	2340
Db	2285	ATCAGGTATGGATCAAAAGGAGCTTTTGTAGTGAAGATGATGAAGCAGGAAAGTCTGGA	2344
Qy	2341	AAAAATCATTATGATGATCGGGCTATAGACAAATTTGCTGATCTCGTGGAGGCA	2400
Db	2345	AAAAATCATTATGATGATCGGGCTATAGACAAATTTGCTGATCTCGTGGAGGCA	2404
Qy	2401	GAGGAGCTCTCAGTGGATGATGAGAGGAGAAATGGATTCTTAAGGCTTTCAAGTTGGCT	2460
Db	2405	GAGGAGCTCTCAGTGGATGATGAGAGGAGAAATGGATTCTTAAGGCTTTCAAGTTGGCT	2464
Qy	2461	AATTTTGAATATATAGATGAAATGAGCGAGCAGCATTTAGAGGACACAGAGAGTCTGCTGCT	2520
Db	2465	AATTTTGAATATATAGATGAAATGAGCGAGCAGCATTTAGAGGACACAGAGAGTCTGCTGCT	2524

Qy	2521	GAAAGCAAAATCTTCAGAGGCAATTTCTGATAGAGCAAGTTATTGGGAAGAGTTGTTAAAA	2580
Db	2525	GAAAGCAAAATCTTCAGAGGCAATTTCTGATAGAGCAAGTTATTGGGAAGAGTTGTTAAAA	2584
Qy	2581	GATAAAATTTGAGCTGCACACAGGCTGAGGAGCTTAATGCTCTTCGAAAAAGAGAGAGT	2640
Db	2585	GATAAAATTTGAGCTGCACACAGGCTGAGGAGCTTAATGCTCTTCGAAAAAGAGAGAGT	2644
Qy	2641	CGCAAGCAGTTGGTATCCATTGAAGAGATGATCTTCTGGTTTGGAAAGATGTGAGCTCT	2700
Db	2645	CGCAAGCAGTTGGTATCCATTGAAGAGATGATCTTCTGGTTTGGAAAGATGTGAGCTCT	2704
Qy	2701	GATGGAGATGAAGTTTATGAGCTGAGTCAACAGATGGTGAAGCAGCAGCAGCAGAGATT	2760
Db	2705	GATGGAGATGAAGTTTATGAGCTGAGTCAACAGATGGTGAAGCAGCAGCAGCAGAGATT	2764
Qy	2761	CAGACGGGTGCGACGGCGTACAGAAAGGGTGCAGATAAATTTGGAAACCACTCCGTTG	2820
Db	2765	CAGACGGGTGCGACGGCGTACAGAAAGGGTGCAGATAAATTTGGAAACCACTCCGTTG	2824
Qy	2821	ATGGAAGGTGAGGGAGATCTTTTCAAGAGTACTGGGTTTCAACAGAGTCAAAAGGCCATT	2880
Db	2825	ATGGAAGGTGAGGGAGATCTTTTCAAGAGTACTGGGTTTCAACAGAGTCAAAAGGCCATT	2884
Qy	2881	TTTGTACAGACTTTGATGAGGTATGAGCTGGCAATTTTGAATTTGGAAGAGATTGTTCTCT	2940
Db	2885	TTTGTACAGACTTTGATGAGGTATGAGCTGGCAATTTTGAATTTGGAAGAGATTGTTCTCT	2944
Qy	2941	CGCTTAAAGCAGAAAGACCTTTTGAAGAAATAAATGAATATGGAATACTCTTCTTGAAGCAC	3000
Db	2945	CGCTTAAAGCAGAAAGACCTTTTGAAGAAATAAATGAATATGGAATACTCTTCTTGAAGCAC	3004
Qy	3001	ATTGCTGAAGAAATAGACGAGAAATCTCCAACCTTTTTCAGATGGTGTGCCAAAGGAAGCA	3060
Db	3005	ATTGCTGAAGAAATAGACGAGAAATCTCCAACCTTTTTCAGATGGTGTGCCAAAGGAAGCA	3064
Qy	3061	CTTAGAATAGAGATGTTCTAGTCAAGATTGCTCTTCTGATACTAGTTTCAAGGAGAGGTG	3120
Db	3065	CTTAGAATAGAGATGTTCTAGTCAAGATTGCTCTTCTGATACTAGTTTCAAGGAGAGGTG	3124
Qy	3121	AAATTTGTAGAGATCATCCAGGAAACCTGTTTTCCCTCTCGCATTTCTTGAAGATTCT	3180
Db	3125	AAATTTGTAGAGATCATCCAGGAAACCTGTTTTCCCTCTCGCATTTCTTGAAGATTCT	3184
Qy	3181	CCCGGACTGAGAGTGGAAAAATTTTGAAGAGGAGAAACATGACAAAGATAATGATACGTGCT	3240
Db	3185	CCCGGACTGAGAGTGGAAAAATTTTGAAGAGGAGAAACATGACAAAGATAATGATACGTGCT	3244
Qy	3241	GTTTTAAAGCATGGGTACGACGCGTGGCAAGCTATTGTTGATGACAAAGAGTTGGGGATC	3300
Db	3245	GTTTTAAAGCATGGGTACGACGCGTGGCAAGCTATTGTTGATGACAAAGAGTTGGGGATC	3304
Qy	3301	CAAGACTTATCTGCAAGAAATTTGAATTTCCCTCACAATAGTTTGTCTGCTGCTGAACAA	3360
Db	3305	CAAGACTTATCTGCAAGAAATTTGAATTTCCCTCACAATAGTTTGTCTGCTGCTGAACAA	3364
Qy	3361	GCTGGTTTGCAGGGCAGAAATGGTACTGGGGGCTCTAATCCGGGAGCAGACACTAACAG	3420
Db	3365	GCTGGTTTGCAGGGCAGAAATGGTACTGGGGGCTCTAATCCGGGAGCAGACACTAACAG	3424
Qy	3421	AATCTGGAAGCGTTTATCTGGGAAACAATATGCTTCTGCTGATGGGGCTCAAGTAAAC	3480
Db	3425	AATCTGGAAGCGTTTATCTGGGAAACAATATGCTTCTGCTGATGGGGCTCAAGTAAAC	3484
Qy	3481	TCGATGTTCTATTATCGGGACATGACAGACGCTTTGTTGAGTTTGTGAAAAAGCCAGTT	3540
Db	3485	TCGATGTTCTATTATCGGGACATGACAGACGCTTTGTTGAGTTTGTGAAAAAGCCAGTT	3544
Qy	3541	CTGCTTTTGGAGAGCGGATGAATTAATGATACGAGAGGAATATTATGAGCTTGGTGGC	3600
Db	3545	CTGCTTTTGGAGAGCGGATGAATTAATGATACGAGAGGAATATTATGAGCTTGGTGGC	3604
Qy	3601	TCATCATCTATCCCTACTGAAAGAACAGAGCTGGAACCAAGATCGCTGACACAGTGGGA	3660

```
Db 3605 TCATCATCTATCCCTACTGAAGAACCAAGAGCTGAACCAAGATCGCTGACACAGTGGGA 3664
Qy 3661 GTGAGCTTTATAGAGTTGATGAAATGCTTATGAGTCTTCTTAAGATGATGCTTATC 3720
Db 3665 GTGAGCTTTATAGAGTTGATGAAATGCTTATGAGTCTTCTTAAGATGATGCTTATC 3724
Qy 3721 ACTTCAGAGAAATATATGGGGCTGCTGTTGACAAACCAAGCGCGGTGCGAAATAGCT 3780
Db 3725 ACTTCAGAGAAATATATGGGGCTGCTGTTGACAAACCAAGCGCGGTGCGAAATAGCT 3784
Qy 3781 CAACATTATACAGATGTGCAAACTTCTTATGATGAGAACCGTTCGGGAATCAGTCCAAGCA 3840
Db 3785 CAACATTATACAGATGTGCAAACTTCTTATGATGAGAACCGTTCGGGAATCAGTCCAAGCA 3844
Qy 3841 TATGTAACACCAACCAAGTACCAAGTGAATGAGAGTCTTCTGCTGCACTCAATCT 3900
Db 3845 TATGTAACACCAACCAAGTACCAAGTGAATGAGAGTCTTCTGCTGCACTCAATCT 3904
Qy 3901 ATCAATGTTAATACCAATCTTTCGATTACATCTGATCAATCAATCAATCAATCAATGAA 3960
Db 3905 ATCAATGTTAATACCAATCTTTCGATTACATCTGATCAATCAATCAATCAATGAA 3964
Qy 3961 GACGACCAAGCCAGACCTTAAACATTTGATGATGAGAACGACGCGCGAAGAAACAAA 4020
Db 3965 GACGACCAAGCCAGACCTTAAACATTTGATGATGAGAACGACGCGCGAAGAAACAAA 4024
Qy 4021 CCGTTAAGAGTGGCGTCTGATCTGAATGCTGAGGAGGAGGAGACATTTGCTGAA 4080
Db 4025 CCGTTAAGAGTGGCGTCTGATCTGAATGCTGAGGAGGAGGAGACATTTGCTGAA 4084
Qy 4081 GCTAGTGAAGTGTGATGTTAAATGGAAGAACCCAAAGAACGAGAACCAAGAAC 4140
Db 4085 GCTAGTGAAGTGTGATGTTAAATGGAAGAACCCAAAGAACGAGAACCAAGAAC 4144
Qy 4141 ATGTCGTTGATGATCACTCACTGATTAATCAAGATTC 4177
Db 4145 ATGTCGTTGATGATCACTCACTGATTAATCAAGATTC 4181
```

```
RESULT 2
AF185577 4180 bp mRNA linear PLN 30-NOV-1999
LOCUS Arabidopsis thaliana chromatin remodeling factor CHD3 (PKL) mRNA,
DEFINITION complete cds.
ACCESSION AF185577
VERSION AF185577.1 GI:6478517
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 4180)
Ogas,J., Kaufmann,S., Henderson,J. and Somerville,C.
PICKLE is a CHD3 chromatin-remodeling factor that regulates the
transition from embryonic to vegetative development in Arabidopsis
Proc. Natl. Acad. Sci. U.S.A. 96 (24), 13839-13844 (1999)
JOURNAL 20040637
MEDLINE 10570159
PUBMED
REFERENCE 2 (bases 1 to 4180)
Ogas,J., Kaufmann,S., Henderson,J. and Somerville,C.
Direct Submission
Submitted (12-SEP-1999) Biochemistry, Purdue, 1153 Biochemistry,
West Lafayette, IN 47907, USA
FEATURES
location/Qualifiers
1. 4180
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/map="between mi238 and GP1"
```

```
gene
CDS
1. 4180
/genes="PKL"
25. 4179
/genes="PKL"
/notes="PICKLE"
/codon_start=1
/product="chromatin remodeling factor CHD3"
/protein_id="AAF13875.1"
/db_xref="GI:6478518"
/tranlation="MSSLVERLRIRSDRPFVNVNLDSDDDDFVKKDRTFEQVEAIVR
TDKENACACGEGSTNLVSCNTCTYAFHAKCLVPLKDAVENNRCPCEVCYSLNEDIK
ILDCEMRPTKSSQSSDAEPKFIQVYLWVWKGSLYLHCDSWPEKFOKAYKSNHR
LKRNVNFRQMSFNSSDEDDFAIVPEVTTVDRIILACREDEKYLKELSYDE
CYWESDSTISFQNEIRFRKDVNSRTRSDVHKRNPDRFQDHTPEFLKGLLHY
QLEGNLFRFSQTHVILADEMGLKTIQSIALLASLPEENLPHLVLAPLSTLN
WEREFATAPQNVNVAVFGTAQARVIRHEFEVLSKDOKI KKKKSQIISSEKOKR1
KDFVLTSYEMINLDSAVLKPWKWECHVIDBEGHRLKNKSKLPSSTQYVSNHRLVT
GTPQLNNDELFMHFLDAGKFGSLEEFEDFNOBEQISRLHKMLAPHLRLRVK
KDYKMDPPKELILRVDLSSLOKEYKAI FTFRNYQVLTCKGGAQISLNNIMMELRV
CCHPYMLEGPEVVIHDANAFKQLSCGKLQLLDRMMVKLEQGHRLVLTFTFOHML
DLEEDYCTHKWOYERIDGVKGAERQIRIDRENAKNKFCFLSTRAAGLGINLAT
ADTVIYDSNPHADLOAMARHLGOTNKVMIYRLINRGTIEERMOLTKKMWLE
HLVGLKLTQNTNQELDLIRYKSELEFASDEDEAGKSKIHYDDAAIDKLDRDLV
EAEVSVDEEENGFLKAFKVFANFETIDENEAALAEARVAESKSGASDRASWE
ELLKDFELHQAELNALGKRKSRKOLVSI EEDDLAGLEDVSSDGSSEASSTDE
AAGQVQVQTRPYRRGRDNLEPTFLMEGERSFVLFNFQSRALFVQVTLMRGAGN
FMWKEFVRLKQKTFEINBYGLFLKHAEEIDENSPTFSQGVKPEGLRIDEVLVRI
ALILVQEKVKFVEDHPGKVPVPSRILRFPGLRSGKI WKEHDKIMIRAVLKHGYR
WQALVDKELG1QBELICKELNFPHISLSAEOAGLOGONGSGNSGNPOTNPGSVI
TGNNSADAGVNSPFYTRDMORRLVFEVKKRVLLEKAMNYEAYEYVGLGSSSI
YNEEPAEPK1ADTVGVSFIEVDEMDLGPKTDPITSEIIMGAADVNNQARVEIAQH
YPMCKLLDENARESQAAYVNNQPPSTKVNESFRALKSINGNTNITLSITSDOSKSH
DDTKPDLNNVNMKDTAEETKPLRGVVDLNVVEEENIEAASGSDVDMEEAKEEP
KNMVD"
```

ORIGIN

```
Query Match 99.5%; Score 4156; DB 8; Length 4180;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGAGTAGTTTGGTGGAGAGCTTCCATACGACTCTGATAGGAACACGAGTTTAAACCTA 60
Db 25 ATGAGTAGTTTGGTGGAGAGCTTCCATACGACTCTGATAGGAACACGAGTTTAAACCTA 84
Qy 61 GATGATTCTGATGATGACGACTTCGTTCTTCTTAAAAAGATCGAACCTTTTGACGAGTCGAG 120
Db 85 GATGATTCTGATGATGACGACTTCGTTCTTCTTAAAAAGATCGAACCTTTTGACGAGTCGAG 144
Qy 121 GCTATTGTCAGAACTGATGCGAAGAAATGCAATGTCAGGCTTGTGGGGAAAGTACTAAT 180
Db 145 GCTATTGTCAGAACTGATGCGAAGAAATGCAATGTCAGGCTTGTGGGGAAAGTACTAAT 204
Qy 181 CTTGTAAGCTGCAATACATGCACTTATCGCTTCCATGCTAAATGCTTAGTTTCCACCTCTT 240
Db 205 CTTGTAAGCTGCAATACATGCACTTATCGCTTCCATGCTAAATGCTTAGTTTCCACCTCTT 264
Qy 241 AAAGATCCTTCCGTCGAAAAATTTGGAGATGCGCTGATGTTAGTCTCTCTTAACGAGATA 300
Db 265 AAAGATCCTTCCGTCGAAAAATTTGGAGATGCGCTGATGTTAGTCTCTCTTAACGAGATA 324
Qy 301 GATAAGATATTGGAATGTGAAATGCGCTCTTACAAAATCTAGTGAACAAGGTTCTCCGAT 360
Db 325 GATAAGATATTGGAATGTGAAATGCGCTCTTACAAAATCTAGTGAACAAGGTTCTCCGAT 384
Qy 361 GCGGAACCGAAGCAAAATTTTGTGAACAGTATCTCGTGAAGTGAAGGAGATTATCATAC 420
Db 385 GCGGAACCGAAGCAAAATTTTGTGAACAGTATCTCGTGAAGTGAAGGAGATTATCATAC 444
Qy 421 CTTCACTGCTCTTGGGTGCTCGAGAGAGGTTCCAGAAGGCTTATAAGTCAAAATCATCGT 480
Db 445 CTTCACTGCTCTTGGGTGCTCGAGAGAGGTTCCAGAAGGCTTATAAGTCAAAATCATCGT 504
Qy 481 TTAAAAACAGAGTGAACAAATTTTCCCGTCAAAATGGAGTCATTTCAATTAACAGGAAGAT 540
```


|||||
505 TTAACACGAGTGAACAATTTTACCGTCAAAATGAGTCAATTAATAACGGAAGAT 564
QY
541 GATTTGTGTCACACGTCCTGAGTGACCACTGTTGATCGGATCTTGGCTGCGAGAG 600
Db
565 GATTTGTGTCACACGTCCTGAGTGACCACTGTTGATCGGATCTTGGCTGCGAGAG 624
QY
601 GAAGATGGGAGCTGGAAATATCTTGTCAATATATAAGAGCTATCTATGATGATGTTAT 660
Db
625 GAAGATGGGAGCTGGAAATATCTTGTCAATATATAAGAGCTATCTATGATGATGTTAT 684
QY
661 TGGAGTCAAGATCAGACATCTCAACCTTCAGAAATGAATTCAGAGTTCAAGAGTGA 720
Db
685 TGGAGTCAAGATCAGACATCTCAACCTTCAGAAATGAATTCAGAGTTCAAGAGTGA 744
QY
721 AATTCTAGAACTCGCAGAAGTAAAGATGTTGACCATAAAAAGAAATCCAGAGACTTTCAA 780
Db
745 AATTCTAGAACTCGCAGAAGTAAAGATGTTGACCATAAAAAGAAATCCAGAGACTTTCAA 804
QY
781 CAGTTGATCATACTCTGAATTCCTCAAGGCTTGTATCATCATACAGCTTGAGGA 840
Db
805 CAGTTGATCATACTCTGAATTCCTCAAGGCTTGTATCATCATACAGCTTGAGGA 864
QY
841 CTTAATTTTTGCGTTCTCGTGTCAAAACAGACGATGTAATCTTCTGCTGATGAATG 900
Db
865 CTTAATTTTTGCGTTCTCGTGTCAAAACAGACGATGTAATCTTCTGCTGATGAATG 924
QY
901 GGACTAGGCAAGACAAATTCAGAGCATTTGCCCTTTAGCTTCACCTTTTGAGGAGAACCTC 960
Db
925 GGACTAGGCAAGACAAATTCAGAGCATTTGCCCTTTAGCTTCACCTTTTGAGGAGAACCTC 984
QY
961 ATTCCGATTTGGTAATGCTCTCTATCGACTCTGTAACCTGGGAGAGAGATTTGCC 1020
Db
985 ATTCCGATTTGGTAATGCTCTCTATCGACTCTGTAACCTGGGAGAGAGATTTGCC 1044
QY
1021 ACATGGCCCCACAGATGAAGTGTATGTAATTTGGCACTGCGCAAGCTCGAGCAGTT 1080
Db
1045 ACATGGCCCCACAGATGAAGTGTATGTAATTTGGCACTGCGCAAGCTCGAGCAGTT 1104
QY
1081 ATCAGAGAAATGAGTTTACTTATCGAAAGATCAAAAAAGATCAAAAAAGAAATCT 1140
Db
1105 ATCAGAGAAATGAGTTTACTTATCGAAAGATCAAAAAAGATCAAAAAAGAAATCT 1164
QY
1141 GGCACAAATAAGTAGCGAAAGCAAGCAAAAAAGAAATCAAGTTTGTATGTCCTCCACATCG 1200
Db
1165 GGCACAAATAAGTAGCGAAAGCAAGCAAAAAAGAAATCAAGTTTGTATGTCCTCCACATCG 1224
QY
1201 TATCAGATGATCAACCTAGATTCAGCAGTTCTAAACCAATTAAGTGGGAGTGCATGAT 1260
Db
1225 TATCAGATGATCAACCTAGATTCAGCAGTTCTAAACCAATTAAGTGGGAGTGCATGAT 1284
QY
1261 GTTGATGAAGGTCACTGACTGAAAAATAAGGATTCAAAGCTTCTCTCAATTGACACAG 1320
Db
1285 GTTGATGAAGGTCACTGACTGAAAAATAAGGATTCAAAGCTTCTCTCAATTGACACAG 1344
QY
1321 TATTCAAGTAAACACCGTATCTTCTCAGAGGAACACCACTTCAGAAACAACTTGGATGAA 1380
Db
1345 TATTCAAGTAAACACCGTATCTTCTCAGAGGAACACCACTTCAGAAACAACTTGGATGAA 1404
QY
1381 CTTTTCATGCTCATGCAATTTCTGTATGCGGGAGTTTGAAGTTTGGAGGATTCAG 1440
Db
1405 CTTTTCATGCTCATGCAATTTCTGTATGCGGGAGTTTGAAGTTTGGAGGATTCAG 1464
QY
1441 GAGGAGTTCAAGAGATATTAATCAAGAGGAGCAGATCTCAAGGTTGCAAAAAATGTTGGCT 1500
Db
1465 GAGGAGTTCAAGAGATATTAATCAAGAGGAGCAGATCTCAAGGTTGCAAAAAATGTTGGCT 1524
QY
1501 CCACATTTGCTCAGAGGGTAAAAAGACGTAATGAAGACATGCCCCCCCAAAAAGGAG 1560
Db
1525 CCACATTTGCTCAGAGGGTAAAAAGACGTAATGAAGACATGCCCCCCCAAAAAGGAG 1584
QY
1561 CTCAATTTGGTGTGATCTGAGCAGTCTGCGAGAAAGATTAATCAAAAGCTATTTTACC 1620
|||||

1585 CTCAATTTGGTGTGATCTGAGCAGTCTGCAGAAAGAAATATTACAAAGCTATTTTACC 1644
QY
1621 CGTAATTTATCAAGATATTGACAAAAAGGGAGGTGCTCAAAATTTCCCTTAATAACATTATG 1680
Db
1645 CGTAATTTATCAAGATATTGACAAAAAGGGAGGTGCTCAAAATTTCCCTTAATAACATTATG 1704
QY
1681 ATGGAATTAAGAAAGTATGCTGCCATCTTATATGCTAGAGGGTGTGAGCCAGTTATT 1740
Db
1705 ATGGAATTAAGAAAGTATGCTGCCATCTTATATGCTAGAGGGTGTGAGCCAGTTATT 1764
QY
1741 CAGCAGCAAAATGAAGCTTTCAAAACAACTTTTGGAGTCTTGTGAAAAAGCTGCAACTTCTA 1800
Db
1765 CAGCAGCAAAATGAAGCTTTCAAAACAACTTTTGGAGTCTTGTGAAAAAGCTGCAACTTCTA 1824
QY
1801 GATAAAATGATGTCAAACTGAAGAGCAAGGACACAGAGTCTCTTAATATACACAGTTT 1860
Db
1825 GATAAAATGATGTCAAACTGAAGAGCAAGGACACAGAGTCTCTTAATATACACAGTTT 1884
QY
1861 CAGCATATGCTGCACTTACTTGAAGACTACTGTACCCATGAAGAAATGGCAGTAGCAGCGA 1920
Db
1885 CAGCATATGCTGCACTTACTTGAAGACTACTGTACCCATGAAGAAATGGCAGTAGCAGCGA 1944
QY
1921 ATTGATGGAAGGTGCGGAGCTGAGCGCAAAATACGCATAGATCGGTTCAATGCCAAA 1980
Db
1945 ATTGATGGAAGGTGCGGAGCTGAGCGCAAAATACGCATAGATCGGTTCAATGCCAAA 2004
QY
1981 AATTCTAACAAAGTTTGTCTCTCCACAAGAGCTGGTGGCTTTAGGAATAAATCTT 2040
Db
2005 AATTCTAACAAAGTTTGTCTCTCCACAAGAGCTGGTGGCTTTAGGAATAAATCTT 2064
QY
2041 GCAACGGCTGATACAGTAATCATTTATGACAGTGACTGGAATCTCTATGCTGATCTTCAA 2100
Db
2065 GCAACGGCTGATACAGTAATCATTTATGACAGTGACTGGAATCTCTATGCTGATCTTCAA 2124
QY
2101 GCAATGCTAGAGCTCATCGACTTGGCACAACAAATAAGGTGATGTTTATAGGCTCAT 2160
Db
2125 GCAATGCTAGAGCTCATCGACTTGGCACAACAAATAAGGTGATGTTTATAGGCTCAT 2184
QY
2161 AACCGAGGCACCAATTGAAGAAAGATGATGCAATTGACTAAAAAGAAATGGTTCTAGAG 2220
Db
2185 AACCGAGGCACCAATTGAAGAAAGATGATGCAATTGACTAAAAAGAAATGGTTCTAGAG 2244
QY
2221 CATCTTTGTTGGGAAACTCAAAAACAAATTAATCAGGAAGTTAGATGACATC 2280
Db
2245 CATCTTTGTTGGGAAACTCAAAAACAAATTAATCAGGAAGTTAGATGACATC 2304
QY
2281 ATCAGGTTATGATCAAAAGGAGCTTTTGTAGTGAAGATGATGAAGCAGGAAGTCTGGA 2340
Db
2305 ATCAGGTTATGATCAAAAGGAGCTTTTGTAGTGAAGATGATGAAGCAGGAAGTCTGGA 2364
QY
2341 AAAATTCATTTATGATGATGCGCTATAGACAAATTTGCTGATCGTATCTCGTGGAGCA 2400
Db
2365 AAAATTCATTTATGATGATGCGCTATAGACAAATTTGCTGATCGTATCTCGTGGAGCA 2424
QY
2401 GAGGAAGTCTCAGTGAATGATGAAGAGGAGAAATGGATTTCTTAAAGGCTTTCAAGGTGGCT 2460
Db
2425 GAGGAAGTCTCAGTGAATGATGAAGAGGAGAAATGGATTTCTTAAAGGCTTTCAAGGTGGCT 2484
QY
2461 AATTTTGAATATATAGATGAATGAGGAGCAGCATTAGAGGCACAGAGAGTCCGCTGCT 2520
Db
2485 AATTTTGAATATATAGATGAATGAGGAGCAGCATTAGAGGCACAGAGAGTCCGCTGCT 2544
QY
2521 GAAAGCAAAATCTTCAGCAGGCAATTTCTGATAGAGCAAGTTTATTTGGGAAGAGTTGTTAAAA 2580
Db
2545 GAAAGCAAAATCTTCAGCAGGCAATTTCTGATAGAGCAAGTTTATTTGGGAAGAGTTGTTAAAA 2604
QY
2581 GATAAAATTTGAGCTGCACAGGCTGAGGAGCTTTAATGCTCTTTCGAAAAAGGAAGAGT 2640
Db
2605 GATAAAATTTGAGCTGCACAGGCTGAGGAGCTTTAATGCTCTTTCGAAAAAGGAAGAGT 2664
QY
2641 CGCAAGCAGTGTGATCTGAGCAGTCTGCGAGAAAGATGATCTTGTGTTTGGAAAGATGTGAGCTCT 2700
Db
2665 CGCAAGCAGTGTGATCTGAGCAGTCTGCGAGAAAGATGATCTTGTGTTTGGAAAGATGTGAGCTCT 2724

```
Qy 2701 GATGGAGTAAAGCTTATGAAGCTGAGTCAACAGATGCTGAAGCAGCAGCAGCAAGGAGTT 2760
Db |||||||
Qy 2725 GATGGAGATGAAGTTAAGAGCTGAGTCAACAGATGCTGAAGCAGCAGCAAGGAGTT 2784
Db |||||||
Qy 2761 CAGACGGGTGACGGCGGTACAGAAAGAAAGGTCGCGATAAATTTGGAAACCAACTCCGTTG 2820
Db |||||||
Qy 2785 CAGACGGGTGACGGCGGTACAGAAAGAAAGGTCGCGATAAATTTGGAAACCAACTCCGTTG 2844
Db |||||||
Qy 2821 ATGGAAGGTGAGGGAGATCTTTTCAGAGTACTGGGTTTCAACAGAGTCAAGGGCCATT 2880
Db |||||||
Qy 2845 ATGGAAGGTGAGGGAGATCTTTTCAGAGTACTGGGTTTCAACAGAGTCAAGGGCCATT 2904
Db |||||||
Qy 2881 TTTGTACAGACTTTGATGAGTATGGAGCTGGCAATTTTGAATGCAAGGAGTTTGTTCCT 2940
Db |||||||
Qy 2905 TTTGTACAGACTTTGATGAGTATGGAGCTGGCAATTTTGAATGCAAGGAGTTTGTTCCT 2964
Db |||||||
Qy 2941 CGCTTAAAGCAGAAAGACCTTTTGAAGAAATAAATGAATATGGAATACTCTTCTTGAAGCAC 3000
Db |||||||
Qy 2965 CGCTTAAAGCAGAAAGACCTTTTGAAGAAATAAATGAATATGGAATACTCTTCTTGAAGCAC 3024
Db |||||||
Qy 3001 ATTCTGAAGAAATAGACGAGAAATCTCCAACTTTTCAGATGCTGTGCCCAAGGAAGGA 3060
Db |||||||
Qy 3025 ATTCTGAAGAAATAGACGAGAAATCTCCAACTTTTCAGATGCTGTGCCCAAGGAAGGA 3084
Db |||||||
Qy 3061 CTTAGAATAGAAGATGTTCTAGTCAGAAATGCTCTCTGATACTAGTTTCAAGGAGAGGTG 3120
Db |||||||
Qy 3085 CTTAGAATAGAAGATGTTCTAGTCAGAAATGCTCTCTGATACTAGTTTCAAGGAGAGGTG 3144
Db |||||||
Qy 3121 AAATTTGTAGAAGATCATCAGGGAACCTGTTTCCCTCTCGCATTTCTTGAAGATTC 3180
Db |||||||
Qy 3145 AAATTTGTAGAAGATCATCAGGGAACCTGTTTCCCTCTCGCATTTCTTGAAGATTC 3204
Db |||||||
Qy 3181 CCCGACTGAGAAGTGAAAAATTTGGAAGGAGAAATGACACAGATAATGATACGTGCT 3240
Db |||||||
Qy 3205 CCCGACTGAGAAGTGAAAAATTTGGAAGGAGAAATGACACAGATAATGATACGTGCT 3264
Db |||||||
Qy 3241 GTTTTAAAGCATGGTACGACGCTGCAAGTATTGTTGATGACAAAGAGTTGGGGATC 3300
Db |||||||
Qy 3265 GTTTTAAAGCATGGTACGACGCTGCAAGTATTGTTGATGACAAAGAGTTGGGGATC 3324
Db |||||||
Qy 3301 CAAGAGCTTATCTCAAGAAATGAATTTCCCTCACATAAGTTTGTCTGCTGTGAACAA 3360
Db |||||||
Qy 3325 CAAGAGCTTATCTCAAGAAATGAATTTCCCTCACATAAGTTTGTCTGCTGTGAACAA 3384
Db |||||||
Qy 3361 GCTGTTTTCAGGGGCGAGATGTTAGTGGGGCTCTAATCCGGAGACACAGACTAACCG 3420
Db |||||||
Qy 3385 GCTGTTTTCAGGGGCGAGATGTTAGTGGGGCTCTAATCCGGAGACACAGACTAACCG 3444
Db |||||||
Qy 3421 AATCCTGGAAGCGTTATTACTGGGAACAATAATGCTTCTGCTGATGGGCTCAAGTAAC 3480
Db |||||||
Qy 3445 AATCCTGGAAGCGTTATTACTGGGAACAATAATGCTTCTGCTGATGGGCTCAAGTAAC 3504
Db |||||||
Qy 3481 TCGATGTTCTATTATCGGGACATCGAGACGACTTGTGAGTTTGTGAAAAGCGAGTT 3540
Db |||||||
Qy 3505 TCGATGTTCTATTATCGGGACATCGAGACGACTTGTGAGTTTGTGAAAAGCGAGTT 3564
Db |||||||
Qy 3541 CTGCTTTTGGAGAAGGCGATGAATTAATGATACCGAGAGNATTAATGACTTTGGTGGC 3600
Db |||||||
Qy 3565 CTGCTTTTGGAGAAGGCGATGAATTAATGATACCGAGAGNATTAATGACTTTGGTGGC 3624
Db |||||||
Qy 3601 TCATCATCTATCCCTACTGGAAGAACCAAGAGCTGCAACCAAGATCGCTGACACAGTGGGA 3660
Db |||||||
Qy 3625 TCATCATCTATCCCTACTGGAAGAACCAAGAGCTGCAACCAAGATCGCTGACACAGTGGGA 3684
Db |||||||
Qy 3661 GTGAGCTTTATTTGAGGTTGATGATAATGCTTGTGAGACTTCTTAAGACTGATCCTATC 3720
Db |||||||
Qy 3685 GTGAGCTTTATTTGAGGTTGATGATAATGCTTGTGAGACTTCTTAAGACTGATCCTATC 3744
Db |||||||
Qy 3721 ACTTCAGAGAAATATGGGGGCTGCTTTGACAAACCAAGCCGGGTGCAATAGCT 3780
Db |||||||
Qy 3745 ACTTCAGAGAAATATGGGGGCTGCTTTGACAAACCAAGCCGGGTGCAATAGCT 3804
Db |||||||
```

```
Qy 3781 CAACATTATAACAGATGTGCAAACTTTCTTGATGAGAACGCTCGGAATCAGTCCAAGCA 3840
Db |||||||
Qy 3805 CAACATTATAACAGATGTGCAAACTTTCTTGATGAGAACGCTCGGAATCAGTCCAAGCA 3864
Db |||||||
Qy 3841 TATGTAAAACAAACCAACCCAGGTACCAAGGTGAATGAGAGCTTCGTTGCACTCAAAATCT 3900
Db |||||||
Qy 3865 TATGTAAAACAAACCAACCCAGGTACCAAGGTGAATGAGAGCTTCGTTGCACTCAAAATCT 3924
Db |||||||
Qy 3901 ATCAATGGTAAACATTAACACAATCCCTTTTCGATTACATCTGATCAATCCAAGTCACATGAA 3960
Db |||||||
Qy 3925 ATCAATGGTAAACATTAACACAATCCCTTTTCGATTACATCTGATCAATCCAAGTCACATGAA 3984
Db |||||||
Qy 3961 GACGACACCAAGCAGACCTTAAACAATTTGTGATGAAGACACGGCCGAAGAAACAAA 4020
Db |||||||
Qy 3985 GACGACACCAAGCAGACCTTAAACAATTTGTGATGAAGACACGGCCGAAGAAACAAA 4044
Db |||||||
Qy 4021 CCCTTAAGAGGTGGCGTCTGATCTGAATGTTGGTGGAGGAGAGGAACATTTGCTGAA 4080
Db |||||||
Qy 4045 CCCTTAAGAGGTGGCGTCTGATCTGAATGTTGGTGGAGGAGAGGAACATTTGCTGAA 4104
Db |||||||
Qy 4081 GCTAGTGGAGTCTTGTATGTAAAAATGGAAGACCAAAAGAAAGAGCCAAAGAAC 4140
Db |||||||
Qy 4105 GCTAGTGGAGTCTTGTATGTAAAAATGGAAGACCAAAAGAAAGAGCCAAAGAAC 4164
Db |||||||
Qy 4141 ATGTCGTTGATTGAC 4156
Db |||||||
Qy 4165 ATGTCGTTGATTGAC 4180
Db |||||||

RESULT 3
AX507831
LOCUS 3204 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 2526 from Patent WO0216655.
ACCESSION AX507831
VERSION AX507831.1 GI:23389068
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE
1
Harper, J.F., Kreps, J., Wang, X. and Zhu, T.
Stress-regulated genes of plants, transgenic plants containing
same, and methods of use
Patent: WO 0216655-A 2526 28-FEB-2002;
The Scripps Research Institute (US) ; Syngenta Participations AG
(CH)
FEATURES
source
1. .3204
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
ORIGIN
Query Match 20.4%; Score 850.8; DB 6; Length 3204;
Best Local Similarity 61.0%; Pred. No. 3.4e-189;
Matches 1892; Conservative 0; Mismatches 697; Indels 513; Gaps 11;
Qy 280 GTTAGTCTCTTAAACGAGATAGATAATGATTTGGATTTGCAAAATGCGTCTTACAAAATCT 339
Db |||||||
Qy 121 GATTGTCTCTTGGCGAGATTGAAAAGATTTTGGACCGTGAATGGCGCTCTCTACTGCATCT 180
Db |||||||
Qy 340 AGTGAACAGGTTCTCCCGAT---CGGAAACCGAAGCAATTTTGTGAACACAGTATCTC 396
Db |||||||
Qy 181 AACAAATCCGAATTTCTTGTGACACGGGACCAACCCCTAGTTGTGCGAAACAATATCTT 240
Db |||||||
Qy 397 GTGAAGTGAAGGATTATCATACCTTCACTGCTCTTGGGTGCTGTAGAAGAGGAGTTCCAG 456
Db |||||||
Qy 241 GTGAAGTGAAGGGTTATCGTACTGCTGCTCTTGGGTGCCCGAGGAGGAGTTCCAG 300
Db |||||||
Qy 457 AAGCTTATAGTCAAAATCATCGTT-----TAAACCAACAGAGTGAACAAATTTTCCCGT 510
Db |||||||
```


Qy 2665 GAAGATGATCTTGCTGTTGGAGAGATGTGAGCTCTCATGAGGAGTAAAGATTATGAAGCT 2724
|||
Db 2110 GAAGATGATCTTGCTGTTGGAGAGATGTGAGCTCTCATGAGGAGTAAAGATTATGAAGCT 2169
|||
Qy 2725 GAGTCAACA-----GATGTTGAGCAGCAGGACAA 2754
|||
Db 2170 GATTAAAGATGACCTCTGATGAGGAGGAGGAGACGATGAACCCGAGCAGCTCGACG 2229
|||
Qy 2755 GGAGTTC---AGACGGGTGACGCGCTTACAGAGAAAGGCTCCGATAATTTTGAACCA 2811
|||
Db 2230 AGAAGCCGAGGACGGTTCACAGCCCTTACAGAAAAGGCTCCGATAATTTGAGGAA 2289
|||
Qy 2812 ACTCCGTTGATGAGGAGTGGGAGATCTTTCAGAGTACTGGTCTTCAACCGAGTCAA 2871
|||
Db 2290 ATACCTTTGATGAGGAGGAGGAGATATCTCATGCTGCTGGATTTAAACGACTGAG 2349
|||
Qy 2872 AGGCCATTTTGTACAGACTTTCATGAGGTATGAGCTGGCAATTTTGTGTTGAAGGAG 2931
|||
Db 2350 AGGATATATTTTACGTACGTTTAAAGAGGTATGGAATCTTATTTTGAACACATTTGCC 2409
|||
Qy 2932 TTTGTTCTCGCTTAAAGCAGACGACCTTTGAAGAAATAATGAATATGGAATACTCTTC 2991
|||
Db 2410 GAAACCTTACCGATATTTCTACAAATTTTAAAGTTATACGCAATGTT----- 2459
|||
Qy 2992 TTGAAGCACATTTGCTGAAGAAATAGACGAAATTTCTCAACCTTTTCAGATGTTGCC 3051
|||
Db 2460 -----CTATGCAGATGGAGTTTCCC 2478
|||
Qy 3052 AAGGAAGGACTTAGAATAGAGATGTTCTAGTACAGATTGCTCTCTGATACTAGTTTCAG 3111
|||
Db 2479 AAGAAGGGATTAAGCAGTACGAGCTACTAGTACGATGACTTTTATGATGCTAGTAAAG 2538
|||
Qy 3112 GAGAAGGTGAATTTGTAGAGATCATCCAGGGAACCTGTTTCCCTCTCGCATTTCTT 3171
|||
Db 2539 GAGAAGGTGCAATTTTGGACAACTCCGACACACTGTTTT---CTTACTACGTC 2595
|||
Qy 3172 GAAAGATTTCCCGACTGAGAAGTGGAAAAATTTGGAAGGAGGAACATGACAAGATAATG 3231
|||
Db 2596 ATCAGTAAGTACAATTTGAGAAATGGAGCATTTTCTAAGGAAGAACATGATAGTACTG 2655
|||
Qy 3232 ATAGCTGCTGTTTAAAGCATGGTACGAGCGTGGCGAGCTATGTTGATGACAAAGAG 3291
|||
Db 2656 ATTCTGCTGTTTCCAAAGCATGGCTATGTTAGATGGGTGGCGATCGTTTGAAGACGAAG 2715
|||
Qy 3292 TTGGGATCCCAAGAGCTTATCTGCAAGAAATGAATTTCCCT 3333
|||
Db 2716 ATAGGTTCCAAAGAGTTGCTGCAAGACCTGATATCCCT 2757
|||

RESULT 4
BC046866
LOCUS BC046866 6398 bp mRNA linear VRT 04-NOV-2003
DEFINITION Xenopus laevis chromodomain helicase DNA binding protein 4, mRNA
(CDNA clone MGC:52739 IMAGE:4683903), complete cds.
ACCESSION BC046866
VERSION BC046866.1 GI:28422179
KEYWORDS MGC.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 6398)
Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W.
and Richardson, P.
Genetic and genomic tools for Xenopus research: The NIH Xenopus
initiative
Dev. Dyn. 225 (4), 384-391 (2002)
JOURNAL 22341132
MEDLINE 12454917
PUBMED
REFERENCE 2 (bases 1 to 6398)
Strausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullany, S.J., Bosak, S.A., McSwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Schevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Snailus, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16999-16903 (2002)
22388257
12477932
3 (bases 1 to 6398)
Klein, S. and Strausberg, R.
Direct Submission
Submitted (14-FEB-2003) National Institutes of Health, Xenopus Gene
Collection (XGC), National Institute of Child Health and Human
Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD
20892-7510, USA
NIH-XGC Project
Contact: XGC help desk
Email: cgabps-remail.nih.gov
Tissue Procurement: Dr. Igor Dawid
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 94 Row: g Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.
FEATURES
Location/Qualifiers
1..6398
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="MGC:52739 IMAGE:4683903"
/tissue_type="Embryo, stage 31/32, Xenopus"
/clone_lib="NICHX_XGC_Emb4"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
1..6398
/gene="b230399n07-prov"
/note="synonym: MGC52739"
/db_xref="LocusID:380196"
178..5859
/codon_start=1
/product="b230399n07-prov protein"
/protein_id="AAH46866.1"
/db_xref="GI:28422180"
/db_xref="LocusID:380196"
/translation="MASGLGSPSGDGEIVMLNSSIHQDDLLNDELSEVETPKIKK
KKPKKAKSKSKTKVRQKEDDSSGENDYDNDGGPIRDSGSDYTPGKKKK
KMTPKREKSKVRKEDDDDDDDSKPKSKGQLDDWGMEDIDVFTEDDYRTLT
NYKAFSQFVRPLIAAKNPKIAVSKMMVLGAKWREFTNNPLKSGSAGAAAAAV
AVVESVAVTEVIVPPPPPPVVEVPLKAKTKGKGNARKKPKVPPETKTKP
KAKVAPLKI.KLGGFGSKRKSSEEDIDADSDYDSGMNSLSVSGSTKSSRSRK
KUKATKKKKKEBEGAAGADYETDHDQYCEVQQCGEILICDTCPRAYHNVCLDPDM

	DKAPEKWSCPHEKGEVQWEAKEDNSLDDDDLDLDAVGDPEEDHMEFCRCVKCGGG LLCCDVCPSYTHCLNPPEIPNGEWLCPRTCPPLCKGKIQKILITLWKGQPPPTP VSRPADPDPAALPLEGPEREFVFKWHSYLHSCSWTQELQHLQCMVFNHVSQR KNDMDPPAGDYGVDSEKSRKRPDKPKYLEMEKFRYRGIKPEWMMHRLVNLHSDV KGVYHLYIKWRDLVDOASWEAESDIQDYDIYKQYWNHRELMDGDEGRPKGVKK VKRLKDRPDPVADPTVKYDRQPDYLDATGTLHPYQLEGJNLWLFWSAQOCDTIL ADSMGLGKTQVTAFLVSLYKEGSHGKPLVSLSTIINWEREFEPWAPDVMYVYV GDKDSRAVIRENFEPSGNAIRGGKASRWKKEASVKFVHLVLSYELITDITAVLSI FWNLVVDDEAHLRNQSKQFFRLNGYPLQHKLLLTGTPLQNNLELFLHLNPLTPE DNLEGFLEEFADIAKEDQIKLHLDLPHMLRLKADVPKNMPSKTEIIVRVELSPM QKCYKYPILTRNFEALNTRGGGQVSLNVMMDLKKCCNHPYLPFAAMEAPKPNMG YDGSALIKAGKFLQKMLKLDGHRVLIIFSQMTKMLDLEDFMEHGEYKVERID GGLTGNMROAIDRENAPOAQOFCFLSTFAGSLGILNATADTVIYDSDWPHNDIQ AFSRAHRIQNRKMYIRFVTRASVEERITQVAKKMMHLVVRGLSGTSMKSKQ ELDDILKFGFEELKMFATFEGENKEGEDISVTHYDDKAIARLRLNDQETEPLEQL GNNEYLSFKAQYVVEEMGEDEEVREIIRKQESVDPDYWEKLLRHHEQOQDL ARNLHGKRIKQVYNDGSOEDRDQODDSQSVASBEGDDFERSBAARP NRGLKNDKDKPLPILLARVGGNIIVLGFNAROKAFILNAIMEYGMPPDQAFITOWL RDLRGSEKFKAYVSLFMRHLCEPDAGATFADGVPRGLSRQHVLRIGVMSLIR KVOEFHEVNGRMPLELAENKWSKIDSPKTPSTSPGDTQPNTPAPPASAE DSTKSEVPVKEETDNEVKPEENNETTOPTDQSPAAQTEKEEPEVQBERTSP PPPPPPPEMETDQPEPEKDEVITLDDKDDVQEVTLQNGTTPKTEABEKTCK ASAAAKORFMFIADGFTLHSLWNEBERAATVTKTVEIWHRRHDYLLSGIITH GYARMODIONDVYALINEPFGKENVNGFLEIKNKLARRFKLLQALVIEROLRA AYUNMSDDPSHSMALNTRFAVECLAESHQHSKESMAGNKPNANVHLKVLQLESL LSDMKADVTRLPATARIIPVAVRLQMSERNILSRASRGEPQPOQITQQ" 1264. .1401																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
--	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Db 3515 TCCTCTGTCTACTCGTGTGGAGGTTTAGGAATTAATCTGGCAATGCAGATCTAGT 3574
Qy 2060 TCATTTATGACAGTGAAGTCTCTCATCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
Db 3575 TCATATACGACTCCGACTGGAATCCTCACAATGATATTTCAGGCTTTTCAGCAGGCTCATC 3634
Qy 2120 GACTTGTCCAAACAAATAAGTGTATGATTTATAGGCTCATAAACCCAGGACCAATTTGAAG 2179
Db 3635 GTATTGTGCAGACAGAAAGGTGATGATCTATCGATTGTGACCCGTGATCTGTAGAGG 3694
Qy 2180 AAAGGATGATGCAATTAACAAAGAAATGTTCTAGAGCATCTTTGTTGGG---- 2235
Db 3695 AGAATCACTCAGGTTGCCAAAGAAAGATGATGCTTACTCATCTCTGTGTACGGCTG 3754
Qy 2236 -----AACTCAAAACAAACATTAATCAGNAGAGTTAGATGACATCATCAGGTATG 2290
Db 3755 GCCTTGTGATCAAAACATGGCTCCATGTCCTCAACAGGAGCTTGATGATATCCTCAAGTTG 3814
Qy 2291 GATCAAAAGGAGCTTTTGTCTAGTGAAGATGATGAAGCAGGAAGTCTGGAAAAATTCATT 2350
Db 3815 GAACAGAGGAGCTTTTAAAGATGAAGCTACAGNAGGTGTGAAAAAAGAGGAGAGG 3874
Qy 2351 ATGATGATG-----CGGCTATAGACAAATTTGCTTGATCGTATCTCGTGAGCGAGG 2404
Db 3875 ATATCAGTGTCTATCCATTACGATGACAAAGGCCATTGCTGCTTATTAGACCGGAACCGG 3934
Qy 2405 AAGTCTCAGTGGATGATGAAGAGAGATGATGATCTTTAAAGGCTTTCAAGTGGCTTAATT 2464
Db 3935 ATGAGACTGAAGAACCAAGAAATTCGAATTTGGGAATGAAATGATCTTTAGTCTCTTCAAGG 3994
Qy 2465 TTGAATATATAGATGAATAATGAGCAGCAGCATTAGAGGACACAGAGTCTGCTGCTGAAA 2524
Db 3995 TGGCACAGTATGTGTCGCTGAAGAGGAGATGGGGATGAAGAGAGTGTGCGTGAGA 4054
Qy 2525 GCAAAATCTCAGCAGGCAATTTCTGATAGACAAAGTTATTGGGAAGAGTTGTTAAAAAGATA 2584
Db 4055 TTATTAACAAGAAAGAGTGTAGAT---CTGATTACTGGGAGAAATCTTCAGGCATC 4111
Qy 2585 AATTGAGCTGCACCGCTGAGAGCTTAATGCTCTTGGAAAAAGAGAGTGCCTCA 2644
Db 4112 ACTATGAGCAGCAGCAGGAGACCTGCGCCGAAATCTAGGCAAGGAAGCGAATACGCA 4171
Qy 2645 AGCAGTTGGTATCCATTGAAG-----AAGATGATCTTGTGTTTGGAAAGTGTAGCT 2698
Db 4172 AGCAGTGAATCAATGATGGCAGCCAGAGGACAGAGATTCGCAAGATGATCAGTCTG 4231
Qy 2699 CTGATGAGATGAAGTTATGAAGCTGAGTCAACAGATGTTGAAGCAGCAGGACAGGAG 2758
Db 4232 ATATCAGTCGGATTACTCTGTGCTTCTGAAGAAAGGAGATGAAGATTTTGTATGAGAGAT 4291
Qy 2759 TTCAGCGGTCGACGCGCTGACAGAAAGGTCGCGATATTTGGAA-----CCAA 2812
Db 4292 CAGAAGCTGCTGTCGCGCCAAACAGGAAGGTTTGGCGAATGACAAAGACAAAGCCGCTTC 4351
Qy 2813 CTCCTGTTGATGGAAGGTGAGGGAGATCTTTACAGATCTAGGTTTCAACACAGAGTCAAA 2872
Db 4352 CTCCTCTACTGGCTGTTGTTGGTGCATATAGAGTGTCTGATTCATGATCCCGGCAAA 4411
Qy 2873 GGGCCATTTTGTACAGACTTTGATGAGGTATGAGGCTGCGCAATTTTGTATGGA----- 2926
Db 4412 GAAAGGCTTTCCTGAATGCCATCATGCTGATGGATGCGCCCTCAAGATGCTTTTACCA 4471
Qy 2927 ---AGGAGTTGTTCTCGCTTAAGCAGAGAGCTTTGAAGAAATAAATGAATATGAA 2983
Db 4472 CTCAGTGGCTTGTACAGAGATCTCGGTGGAATACTGAGAAAGAAATCAAGGCTTATGAT 4531
Qy 2984 TACTCTTCTTGAAGCACAATGCTGAAGAAATAGACGAGAAATCTTCCAAACCTTTTCAGATG 3043
Db 4532 CCCTCTTCATGCGGCACCTCTGTGAACCTGGGCGAGATGAGCGGAGACTTTTGTCTATG 4591
Qy 3044 GTGTGCCCAAGGAAGGACTTAGAATAGAAATGTTCTAGTACAGAAATTTGCTCTTCTGATAC 3103
Db 4592 GAGTCCCGAGAGAGAGGTCTTCTCGCGCAGCATGTGCTTAAACAAGAAATTTGGGCTCATGTCTC 4651

Qy 3104 TAGTTTCAGGAGAGGT 3119
Db 4652 TGATTCGCAAAAGGT 4667

RESULT 5

BC060721
LOCUS 4343 bp mRNA linear ROD 25-NOV-2003
DEFINITION Mus musculus cDNA clone IMAGE:6406263, partial cds.

ACCESSION BC060721

VERSION BC060721.1 GI:38511411

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 4343)
Klausner, R.D., Collins, F.S., Wagner, L., Shemmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.D., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.C., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

TITLE

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL MEDLINE

PUBMED 12477932

REFERENCE 2 (bases 1 to 4343)

Strausberg, R.

Direct Submission

Submitted (03-NOV-2003) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabs-r@mail.nih.gov

Tissue Procurement: Dr. Jim Lin, University of Iowa

cDNA Library Preparation: M. Bento Soares, University of Iowa

Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) mcdpaxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,

R. M.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAC Plate: 125 Row: f Column: 19

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis.

FEATURES

source

1. 4343 Location/Qualifiers

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

```
/clone="IMAGE:6406263"
/tissue_type="Brain, enriched mouse brain 12.5dp"
/clone_lib="NIH_BMAP_F00"
/lab_host="DH10B"
/note="Vector: pYX-ASC"

ORIGIN

Query Match      10.4%; Score 436.4; DB 10; Length 4343;
Best Local Similarity 57.2%; Pred. No. 1.3e-91;
Matches 995; Conservative 0; Mismatches 691; Indels 54; Gaps 9;

QY 811 GGCTGTATCATCATACAGCTGAGGGACTTAATTTTTCGGTCTCGTCTCAAAA 870
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1573 GGCACATGACATGATATACGTGGAAGGGTGAATGGCTGCTCGTGGGCCAA 1632
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 871 CAGACGATGATATCCCTTGTGTGAATAGGACTAGGCAAGACAATTCAAAGCATGGCC 930
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1633 GGCACATGACATGATATCCCTTGTGTGAATAGGACTAGGCAAGACAATTCAAAGCATGGCC 1692
QY 931 CTTTATAGCTTCACTTTTGGAGGAACTCA-----TTCCGCAATTTGGTAATGTCTCT 984
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1693 TTTCTCTACTCACTGTATAAGGAGGGCCACACAAAGGTCCTTCCCTGGTGAAGTCTCCA 1752
QY 985 CTATCGACTCTGGTAACTGGAGAGAGATTTGCCACATGGGCCCCACAGATGAACGTG 1044
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1753 CTCTCCACCATCATCAACTGGGAACGGGAGTTCAGATGTGGGCACCCCAAGTTCTATGTG 1812
QY 1045 GTTATGTATTTTGGCACTGGCAAGCTCGAGCAGTTTATCAGAGAACATGATTTTACTTA 1104
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1813 GTACATATAGGGCGGACAGGACAGCCGGGCCCATATAGAGAAACGAGTTTTCGTTT 1872
QY 1105 TCGAAAGATCAAAAAAGATCAAGAAAAAGAAATCTGGACAAATTAAGTAGCGAAAGCAAG 1164
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1873 GAGGATAATGCCATAAAGCGGG---AAGAAAGCTTTCAAGATGAAGAGAGAGGCACAG 1929
QY 1165 CAAAAAGATCAAGTTTGTATGTCCTCTCATCTGATGATGATGATCAACCTAGATCA 1224
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1930 -----GTGAAGTTCCATGCTCTCTGACATCATACGAGCTGATCACCATCGATCAG 1980
QY 1225 GCAGTTCTTAAACCAATTAAGTGGAGTGCATGATTTGATGAAGTGCATCGACTGAAA 1284
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1981 GCGGCGCTTGGCTCCATCCGCTGGGCTGCTGTTGTGGTGAAGGCTCATCGACTCAG 2040
QY 1285 AATAAGGATTCAGAGCTGTTCTCTTCAATGACACAGATTAATCAAGTAACCCGCTATCTT 1344
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2041 AACACACAGTCCAAAGTTTTCAGGGTCTCAATGGCTACAAGATAGATCATAGTTGTTG 2100
QY 1345 CTGACAGGAACACCATTTTCAGAACACTTGGATGAACCTTTTCATGCTCATGCTATTTCTT 1404
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2101 CTGACGGGACCCCGCTGCAGAACCACTGGAGGAGCTCTTCCACCTTCTCAACTTCCCTC 2160
QY 1405 GATCGGGGAAGTTTGAAGTTTGGAGGTTTCCAGGAGGTTCCAGAGATTAATCAAA 1464
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2161 ACCCCAGAGAGTTTCAACAACTTGGAGGGCTCTTGGAGAGTTTGGCCACATATCAAAA 2220
QY 1465 GAGGAGAGATCTCAAGGTTGCACAAAATGTTGGCTCCACATTTGCTCAGAAGGTAATAA 1524
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2221 GAGGACAGATTGAAGAACTGCATGATTTACTGGGCGCCACATATGCTTCGGAGACTCAAG 2280
QY 1525 AAGACGATTAAGAAACATGCCCCCCCCCAAAAGAGGCTCAATTTTGGTGTGATCTGAGC 1584
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2281 GCGGATGCTTTAAGAACATGCCACCAAGCAGCAGCTCATTTGTTCCAGTGGAGCTAAGC 2340
QY 1585 AGTCTGAGAGAGATTAATCAAGCTATTTTACCCTGTAATTAATCAAGTATTGA---CA 1641
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2341 CCCATGAGAGAAATACTACAGTACATCTTAATCGAAATTTTGGGCTTGAATCT 2400
QY 1642 AAAAAGGAGGTGCTCAAAATTTCCCTTAATAACATTAATGATGAATTAACGAAAGTATGC 1701
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2401 CGAGTGGTGGAAACCAAGTGTCTCTGCTTAACATCATGATGACCTTAAGAGTGTCTGC 2460
QY 1702 TGCCATCCTTA---TATGCTAGAGGGTGTGTAGCCAGTATTTCACGACGCAATTAAGCT 1758
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
DB 2461 AACACCCGTACTCTTCCCGTGGCTGCTATGAGTCTCTCTAACTCCCGAGTGGGCT 2520
QY 1759 TTCAAAC-----AATTTTGGAGTCTTGTGGAAGCTGCAACTTTCTAGATAAATGATG 1812
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2521 TATGAGGGTGGGCGCTTATTAAGTCAATCAGGGAAGCTGCTGCTGCTGCAAGAAGTCTG 2580
QY 1813 GTCAAACTGAAGAGAGAGACACAGAGTCTTAATATACACAGATTTTACGATATCTG 1872
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2581 CGCAAGCTGAAGGAGGAGACACAGAGTGTCTCATCTTCTCCAGATGACCAAAATGTTA 2640
QY 1873 GACTTACTTGAAGACTACTGTACCCATAGAATAAGGATGACGAGCGAATTTGATGGAAG 1932
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2641 GACCTTCTGGAGACTTCTAGACTATGAAGGTATGAATGATGAGGCATCGATGTGGT 2700
QY 1933 GTTGGCGGAGCTGAGGCGCAAAATACGATAGATCGGTTCATGCCCAAAAATTTTAAACAAG 1992
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2701 ATCAGGGGCGCTCAGCGAGGAGGCCATTGATCGCTTCAATGCCCTGTGCCCCAACAG 2760
QY 1993 TTTTGTGTTTGTCTCCACAGAGCTGTGGCTTAGGAATAAATCTTTCGAAACGGCTGAT 2052
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2761 TTCTGCTTCTCTGTCCTCCACCGAGCTGTGGCTGGGCATCAACTTGGCCACTGTCTGAC 2820
QY 2053 ACAGTAATCATTTTATGACAGTGTGAATCTCATGCTCATCTTCAAGCAATGGCTGAGA 2112
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2821 ACTGTCAATCATCTTCGATCTGATCGAACCCCAATAATGACATCCAGGCTTTCAGCAGA 2880
QY 2113 GCTCATCGACTTGGCCAAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGACCC 2172
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2881 GCCATCGGATTTGGCCAGGCAACAAGTATGATTTTACCGGTTTGTGACCCGAGCTCT 2940
QY 2173 ATTGAAGAAGATGATGCAATTTGACTAAAGAAAATGGTTCTTAGAGCATCTTTGTTGTT 2232
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2941 GTGGAAGAAGCAATCAACAGGTGGCTTAAGAGAAGATGATGCTGACACACCTAGTGGT 3000
QY 2233 GGGAAA-----CTCAAAACACAAAACATTAATCAGGAAGTTAGATGATCATCATC 2283
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3001 CGGCCAGGCTGGGCTCCAGGCTGGCTCCATGTCCAAGCAGGAGCTGGACGACATCCTC 3060
QY 2284 AGTATGGATCAAGAGGCTTTTGTGTAGTGAAGATGATGAAGCAGGAAAGTCTG----- 2338
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3061 AATTTGGACCCGAGGAGCTATTCAAGATGAATGAGGAGAGATTAAGGAGGAGAT 3120
QY 2339 ----GAAAAATTCATATGATGATGGGCTATAGACAAATTTGTTGATCTGATCTCGTG 2394
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3121 AGCAGTGTGATCCACTATGACAATGAAGCCATCGCTCGACTGTGGACCGCTAACAGGAT 3180
QY 2395 GAGGACAGGAAGTCTCAGTGGGATGATGAAGAGAGATGATTTCTTAAGGCTTTTCAAG 2454
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3181 GCGACTGAGGA-----CACGGATGTGCAGAACATGAACGAATACCTCAGCTCCTTCAAG 3234
QY 2455 GTGGCTAAATTTTGAATATATAGATGAAATGAGGAGGAGCATTTAGAGGACAGAGAGTC 2514
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3235 GTGGCACATGATGCTGCTGCTGAAGACACAAGATTGAGAGATCGAGCGAGAGATCATC 3294
```

RESULT 6

```
BC058578
LOCUS
DEFINITION Mus musculus chromodomain helicase DNA binding protein 4, mRNA
ACCESSION BC058578
VERSION BC058578.1 GI:35193270
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 6438)
AUTHORS Strausberg R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
```


Query Match 10.28; Score 427.4; DB 10; Length 6438; Best Local Similarity 52.98; Pred. No. 1.7e-89; Matches 1262; Conservative 0; Mismatches 1026; Indels 99; Gaps 12;									
QY	818	TACATCCATACAGCTTGAGGAGCTTAATTTTGGGGTTCTCGTGGTCAAAACAGACGC	877						
DB	2333	TGATCCCTACCAAAATGGAAGGCTTAACCTGGCTCCGCTTCTCTGGGCTCAGGGAACCG	2392						
QY	878	ATGTAATCTCTGTGTGTAATGGGACTAGGCAAGCAATTCAAAGCAATTCGCCCTTTTAG	937						
DB	2393	ATACTATCTTGGCTGTAGATGGCCCTTGGGAAACTGTGCAGACAGCAGTCTTCTCTCT	2452						
QY	938	CTTCACTTTTGGAGGAACTCAAT-----CGGCATTTGGTAATTTGCTCTCTATCGA	991						
DB	2453	ACTCCCTCTATAAGGAGGCTCACTCCAAAGGCCCTTCTTAGTGAGCGCTCCTCTGTCCA	2512						
QY	992	CTCTGGCTTAACCTGGGAGAGAGTTTGCACATGCGGCCCCACACAGATGAACGTGTTATGT	1051						
DB	2513	CCATCATCAACTGGGAGAGAGTTTGAGATGTGGGCTCCTGATATGTATGTGGTAACCT	2572						
QY	1052	ATTTTGGCACTGGCAGGCTCGAGCAGTTATCAGAGAAATGATGTTTACTTATCGAAAG	1111						
DB	2573	ATGTGGGTGACAGGACAGCGGTGCCATCATCGGGAATGATGTTCTCTTTGAAGACA	2632						
QY	1112	ATCAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAGCAAAAA	1171						
DB	2633	ATGCCATTCGTGGTGGCAAGAGCCCTC-----GAGGATGAAGAAAGAGCAT	2680						
QY	1172	GAATCAAGTTTGATGCTCCTCCTCACATCGTATGAGATGATCAACCTAGATTTCACGAGTTC	1231						
DB	2681	CGGTGAATTTCCATGTTTCTGCTGACATCCTATGAGTTGATCACATTTGATGCGCACT	2740						
QY	1232	TAAAAACCAATTAAGTGGGAGTGATGTTGATGAAGTCTATCGACTGAAAAATAAGG	1291						
DB	2741	TGGGTTCTATTGACTGGGCTGCTCATTTGTGGATGAAGCCCATCGGCTGAAGAAACAAC	2800						
QY	1292	ATTCAAGCTGTTCTCTTCAATTGACACAGTATTCAAGTAACCAAGCTATTTCCTGACAG	1351						
DB	2801	AGTCTAAGTTCTTTTCGAGTTTGAATGGTTACTCACTCCAGCACAAAGCTGTGTCTAAGT	2860						
QY	1352	GAACACCACCTTCAGAACAACTTGATGAACATTTTTCATGCTCATGCTATTTCTTGATCGGG	1411						
DB	2861	GAATCCGTTTACAGAACAACTAGAGGAAGTGTTCATCTGCTCACTTCTCACCCCTG	2920						
QY	1412	GGAGTTTGGAGTTTGGAGAGTTTCCAGAGGAGTTTCAAGATATTAAATCAAGAGAGC	1471						
DB	2921	AGAGTTTCCAACTTAGAAGGCTTCTTGAGGAGTTTGGACACATTTGCCAAGAGAGACC	2980						
QY	1472	AGATCTCAAGGTTGCACAAAATGTTGGCTCCACATTTGCTCAGAAAGGTTAAAAAGACG	1531						
DB	2981	AGATTAAAAAATCCACGACATGCTGGGGCTCATATGTTGGCGGGCTCAAAAGCTGACG	3040						
QY	1532	TAATGAAGACATGCCCCCAAAAGAGGCTCATTTTGGGTGTTGATCTGAGCAGTCTGC	1591						
DB	3041	TCITTCAGAAATATGCCATCCAAGACAGAGCTGATTTGGCGGTGGAGTTGAGTCTATGC	3100						
QY	1592	AGAAAGATATTCAAAAGCTATTTTTTACCCTTAATTAATCAAGTATTGA---CAAAAAAGG	1648						
DB	3101	AGAAAGAGTACTACAAGTATATTCTCACGGGGAATTTTCGAAGCACTTAATGCTCGAGGTG	3160						
QY	1649	GAGTGTCTCAAAATTTCCCTTAATAACATTAATGATGGAATTAACGAAGATATGCTGCCATC	1708						
DB	3161	GTGGCAACCGAGTTTCTCTGCTGAACGTGGTGATGAGATCTTAAGAAATGCTGCACACC	3220						
QY	1709	CTTA-----TATGCTAGAGGCTGTGAGCCAGTTATTTCACGACGCAAAATGAAGCTT	1759						
DB	3221	CTTATCTCTTCCCTGTGGCAGCAATGGAAGCCCTTAAGATGCCCTTAATGGTATGTATGATG	3280						
QY	1760	TCAAACAACTTTTGGAGTCTTGTGGAAGCTGCAACTCTTAGATAAAATGATGCTCAAC	1819						
DB	3281	GCAGTGCCCTAATCAGAGCATCTGGGAAGTTGTTGCTGCTGCAGAAAGATCTTAAGAAC	3340						

QY	1820	TGAAAGAGCAAGCAGACAGAGCTCTAATATACACACAGTTTCAGCATATGCTGCACTTAC	1879
DB	3341	TGAAAGAGGAGGAGGCTGCTGCTCATCTTCTCCAGATGACCAAGATGTTGAGACTGC	3400
QY	1880	TTGAAGACTACTGTATCCCATAAAGAAATGGCAGTACAGAGCAATTTGATGGAAGGTTGGCG	1939
DB	3401	TAGAGGATTTTATAGAACATGAAGGTTATAAATATGAACGTATTGATGGTGAATCACTG	3460
QY	1940	GAGCTAGGCGGCAATATACGATAGATCGGTTCAATGCCAAATAATTTCTAACAGTTTGT	1999
DB	3461	GGAAACATGGCTCAGGAGGCAATTGATCGCTTCAATGCACCGGCTTCAACAGTTCTGCT	3520
QY	2000	TTTTGCTCTCCCAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA	2059
DB	3521	TCCTGTTCTCCACTCGAGCTGGGGCTTGGGATCAATCTGGCCACTCGGACACAGTTA	3580
QY	2060	TCATTTATGACAGTGAATCTCATCTGATCTTCAAGCAATGGCTAGAGCTCATC	2119
DB	3581	TTATATATGACTCTGACTGGAACCCCAATAATGACATCCAGGCTTTAGCAGAGCCACC	3640
QY	2120	GACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCTATAACCGAGGCACCATTTGAAG	2179
DB	3641	GTATTGGGCAAAATAAGAGGTGATGATCTACCGGTTTGTGACCCGTCATCAGTGAGG	3700
QY	2180	AAAGGATGATGCAATTTGACTTAAAGAAATGGTTCTAGAGCATCTTGTGTTGG-----	2234
DB	3701	AGCGCATACGCGAGTGGCAAGAGATGATGCTGACACATTTAGTGGTACGGCTG	3760
QY	2235	----GAAACTCAAAACACAAACATTAATCAGAGAGATTTAGATGACATCATCAGTATG	2290
DB	3761	GCCTGGGCTCCAAGACAGGCTCAATGCTTAAACAGGAGCTTGTATGATATCTCTCAAAATTTG	3820
QY	2291	GATCAAGAGAGCTTTTGTAGTGAAG-----ATGATGAAGCAG	2329
DB	3821	GCATCAGGAGCTATTCAAGGATGAAGCCAGGATGGAGGAGAGACAAAGAGGAGG	3880
QY	2330	GAAGCTCTGGAATAATTTTATGATGATGATGATGATGATGATGATGATGATGATGATG	2389
DB	3881	AGACAGAGTGTATTCCACTATGACGATAAGGCCATTTGAACGCTGCTGGATCGAANACC	3940
QY	2390	TCGTGAGGAGGAGGAGGAGTCTCAGTGGATGATCAAGAGGAGAGATGATTTCTTAAAGCTT	2449
DB	3941	AGATGAGACTGAAGACACAGATTTG-----CAGGCGATGATGATGATTTTGAAGTCTAT	3994
QY	2450	TCAAGTGGCTAAATTTTGAATATATAGATGAAAAATGAGGAGCAGCATTTAGAGCACA	2509
DB	3995	TCAAGTGGCTCAGTATGTGTGTACGGGAAGAGATGG-----GGGAGGAAGAGG	4045
QY	2510	GAGTCTGCTGAAGCAATCTTACAGAGCAATTTCTGATAGCAAGTATTATTTGGGAAG	2569
DB	4046	AGGTAGAACCGGGAATCATAAACAGGAAGAAAGTGTGGATCTCTGAC---TACTGGGAGA	4102
QY	2570	AGTTGTTTAAAGATAAATTTGAGCTGCACCGAGCTGAGGAGCTTAATGCTCTTTGGAAAA	2629
DB	4103	AATGCTGGGACCATTTATGACGAGCAGCAAGATCTAGCCCGAATCTGGGCAAG	4162
QY	2630	GGAGAGAGTTCGCAAGCAGTGTGTTATCCATTCGAAG---AAGATGATTTTGTGTTGG	2686
DB	4163	GAAGAGATCCGTAAACAGGTCAACTACAATGATGGCTCACAGGAGGACCGAGATGGC	4222
QY	2687	AGATGTGAGCTCTGATGGAGATGAAGTTTGAAGCTGAGTCAACAGATGGTGAAGCAG	2746
DB	4223	AGGACACCACTGTCGACCAACCACTCCGATTTACTCAGTGGCTCTCAGAGGAAGGTGATGAAG	4282
QY	2747	CAGGACAGGAGTTACAGCGGTCGAGCGCTACAGAGAAAGGG-----TCGCG	2797
DB	4283	ACTTTGACGAGCGGTGAGAGCTCCCGCAGGCCCATGTCGCAAGGCGCTCGGAATGATA	4342
QY	2798	ATAATTTGGAACCAACTCCGTTTATGAGAGGTGAGGAGATCTTTTCAAGATGCTGGTT	2857
DB	4343	AGATGAAGCATTTACCTCTCTCTGTGGCCGCTGTGTGGGAATATTGAAGTACTTGGTT	4402
QY	2858	TCAACCAAGAGTCAAGGGCCATTTTGTACAGACTTTTGTATGAGGTATGG-----AG	2908

Db 4403 TTAATGCTCGTCAGCGAAAGCAATTTCTTAATGCAATTAATGCGATATGGCATGCCACCTC 4462
 Qy 2909 CTGGCAATTTTGAATGGAAGAGTTTGTCTCTGCTTAAAGCAGAGAACCTTTTGAAGAAA 2968
 Db 4463 AGGATGCTTTTTTACCACTCAGTGGGCTGTGAGAGATCTTCAGGAGCAAGCTCAGAGAAAGAA 4522
 Qy 2969 TAAATGAATATGGAATACTCTCTTGAAGCAGACATGCTGAAGAAATAGACAGAAATTTCTC 3028
 Db 4523 TTAAGGCTTAATGTCATCTTCTATGCGACATTTTGTGTGAGCCCTGGGGCAGATGGGGCTG 4582
 Qy 3029 CAACCTTTTCCAGATGGTGTGCCCAAGGAAGCACTTGAATAGAGATGTTCTTAGTCAGAA 3088
 Db 4583 AGACCTTTGTGTGATGGGCTGCCACGAGAGGCGCTGCTCTGCTCAACATGTTCTCACTAGGA 4642
 Qy 3089 TTGCTCTCTGATACTAGTTTCAGGAGAGAGGTGAAATTTGTAGAGAT 3135
 Db 4643 TCGGTGTCATGTCCTCTGATTCGAAGAAGGTTTCAGGAGTTTGAACAT 4689

RESULT 7

AK128741
 LOCUS 6434 bp mRNA linear PRI 09-SEP-2003
 DEFINITION Homo sapiens cDNA FLJ16830 fis, clone UTERU3022536, highly similar to Chromodomain helicase-DNA-binding protein 4.
 ACCESSION AK128741
 VERSION 1.1 GI:34536266
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 Kanehori, K., Ishibashi, T., Chiba, Y., Fujimori, K., Hiraoka, S., Tanai, H., Watanabe, S., Ishida, S., Ono, Y., Hotuta, T., Watanabe, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A., Oshina, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuno, Y., Nagai, K. and Isogai, T.
 NEDO human cDNA sequencing project
 Unpublished
 2 (bases 1 to 6434)
 Isogai, T. and Yamamoto, J.

TITLE

NEDO human cDNA sequencing project
 Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: Reverse Proteomics Research Institute, HRI and RAB.

COMMENT

Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: Reverse Proteomics Research Institute, HRI and RAB.

FEATURES

source

1. 6434
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UTERU3022536"
 /tissue_type="uterus"
 /clone_lib="UTERU3"
 /note="cloning vector: pME18SFL3"

ORIGIN

Query Match 10.1%; Score 423; DB 9; Length 6434;
 Best Local Similarity 52.0%; Pred. No. 1.8e-88;
 Matches 1237; Conservative 0; Mismatches 1075; Indels 69; Gaps 10;

Qy 818 TACATCCATACACAGCTTGAGGAGCTTAATTTTTTGGGGTCTCGTGGTCAAAAACAGACGC 877
 Db 2279 TGCACCCCTATCAATGAGGGGCTGNAATTTGGTTGGCTTCTCTGGGCTCAGGGCACTG 2338
 Qy 878 ATGTAATCTTGTCTGATGAATAAGGACTAGGCAAGCAATTTCAAAGCAATTTGCCCTTTTAG 937
 Db 2339 ACCCATCTTGGCTGATGATGAGTGGGCTTGGGAAAACTGTACAGACAGCAGTCTTCTCTGT 2398
 Qy 938 CTTCACTTTTGGAGGAG-----AACTCATTCGGCATTTGGTAATTTGCTCCTCTATCGA 991
 Db 2399 ATTCCCTTTTACAAGAGGGGTCATTTCAAAGGCCCTTCTTAGTGAGGGCCCTCTTTCTA 2458
 Qy 992 CTCTGGCTAACTGGGAGAGAGAGTTTGGCCACATGGGGCCCCACAGATGAACGCTGTTATGT 1051
 Db 2459 CCATCATCACTGGGAGCGGAGTTTGAATTTGGGCTCCAGACATGATGTGTAACCT 2518
 Qy 1052 ATTTTGGCACTGCGCAAGCTCGAGCAGTTATTCAGAGAAACATGAGTTTTTACTTATCGAAAG 1111
 Db 2519 ATGTGGGTGACAAAGGACAGCGGTGCCATCATCCGAGAGATGAGTTCTCTTTTGAAGACA 2578
 Qy 1112 ATCAAAAAGATCAAGAAAAGAAATCTGGACAAATAGTAGGAAAGCAAGCAAAAA 1171
 Db 2579 ATGCCATTCGTGTGGCAAGAGGCTC-----CCGCATGAAGAAAGAGGCAT 2626
 Qy 1172 GAATCAAGTTTGTATGTCCTCCTCACATCGTATGAGATGATCAACCTAGATTCAGCAGTTC 1231
 Db 2627 CTGTGAATTCATGTGCTGCTGACATCCTATGAATTTGATCACCATTGACATGGCTATTT 2686
 Qy 1232 TAAAAACAATTAAGTGGGAGTCATGATTTTGGATGAAGGTCTACGATGAAAAAATAAGG 1291
 Db 2687 TGGGCTCTATTGATTGGGCTGCTCATCTGTGATGAAGCCCATCGCTGAGAAACAATC 2746
 Qy 1292 ATTCAAAGCTGTTCTCTTCAATTCACACAGATTTCAAGTAAACCGTATTTCTCTGACAG 1351
 Db 2747 AGTCTAAGTTCTTCCGGGTATTGAATGGTTACTCATCCAGCACAAGCTGTTGCTGACTG 2806
 Qy 1352 GAACACCACTTCAGAAACAACCTTGGATGAACCTTTTCATGTCTCATGCAATTTTCTGATCGG 1411
 Db 2807 GGACACCAATTACAACAACATCTGGAGAGTTGTTTTCATCTGCTCACTTCTCACCCCG 2866
 Qy 1412 GGAAGTTTGAAGTTTGGAGGAGTTCCAGAGAGAGTTCAAAGATATTAATCAAGAGAGC 1471
 Db 2867 AGAGGTTTCCACAATTTGGAAGGTTTGTGGAGGAGTTTGTGTCATTTGCCAAGAGGACC 2926
 Qy 1472 AGATCTCAGGTTGCAAAAATGTTGGCTCCCAATTTGCTCAGAAGGTAAAAAGACG 1531
 Db 2927 AGATAAAAAAAGTGCATGACATCTCTGGGCGGCACATGTTGGCGGCTCAAGACCCATG 2986
 Qy 1532 TAATGAAAGACATGCCCCCAAAAAGAGGCTCAATTTTGGCTGTGATCTGAGCAGTCTGC 1591
 Db 2987 TGTTCAGAACATGCCCTCCAGGACAGACTAATTTGGCTGTGGAGCTGAGCCCTATGC 3046
 Qy 1592 AGAAGAATATTAACAAGCTATTTTACCCGTAATTTATCAAGTATTGA----CAAAAAGG 1648
 Db 3047 AGAAGAATACTACAAGATACATCTCACTCGAAATTTGAAGCACTCAATTCGCCGAGGTG 3106
 Qy 1649 GAGGTGCTCAAAATTTCCCTTAATAACATATATGATGGAATTTACGAAGATATGTCGCCATC 1708
 Db 3107 GTGGCAACCCAGGTGTCCTGCTGAATGTGGTGTGATGATTTAAGAAGTGTCTGAACCATC 3166
 Qy 1709 CTTA-----TATGCTAGAGGGTGTGAGCGAGTATTATTCAGCGCAAAATGAAGCTT 1759
 Db 3167 CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCCTAAGATGCCCTAATGGCATGTATGATG 3226
 Qy 1760 TCAAAACAACCTTTTGGAGTCTTGTGGAAAGCTGCAACTTCTTAGATAAATGATGGTCAAAC 1819
 Db 3227 GCAGTGCCTTAATCAGAGCATCTGGGAAATTAATGTGCTGTCAGAAAAATGCTCAAGAACC 3286
 Qy 1820 TGAAGAGCAAGCAGCAGAGTCTTAATATACACAGATTTTACGATATCTCGACTTAC 1879
 Db 3287 TTAAGGAGGGTGGGCAATCGGTGATCTCATCTTTTCCCAAGTGAACAAGATCTAGACCTGC 3346
 Qy 1880 TTGAAGACTACTGTATCCCAATAAGAAATGGCAGTACGAGCGAATTTGATGGAAAGGTTGGCG 1939

```

Db      3347 TAGAGGATTTCTTGGAACTGAAGGTTATAAATACGAACGATCGATGGTGGAAATCACTG 3406
Qy      1940 GAGCTGAGCGGCAAAATACGATAGATCGGTTCAATGCCAAAATTTCTAAAGTTTGT 1999
Db      3407 GGAACATGCGCAAGAGCCATTTGACCGCTTCAATGACCGGGTCTCAGAGTTCTGCT 3466
Qy      2000 TTTTGTCTTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGAACGGCTGATACAGTAA 2059
Db      3467 TCTTGTCTTCCATCGAGCTGGGGCTTTGGAATCAATCTGGCCACTGCTGACACAGTTA 3526
Qy      2060 TCATTTATGACAGTGAAGTCACTGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
Db      3527 TTATCTATGACTCTGACTGGAACCCCATTAATGACATTCAGGCTTTAGCAGAGCTCAC 3586
Qy      2120 GACTTGGCCAAAACAAATAGCTGATGATTTATAGGCTCATAAACCGAGGCAACCATTTGAAG 2179
Db      3587 GGATTGGCAAAATAAAGAGTAATGATCTACCGGTTGTGACCGCTCAGTGAGG 3646
Qy      2180 AAAGGATGATGAATTAATGATTAAGAAAGAAATGTTCTAGAGCATCTTGTGTTGG----- 2234
Db      3647 AGCGCATACGACAGTGGCAAGAAAGAAATGATGCTGACGCATCTAGTGTGCGGCTG 3706
Qy      2235 -----GAACTCAAAACACAAAATTAATCAGGAAGTTAGATGACATCAGGTATG 2290
Db      3707 GGCTGGGCTCAAGACTGGATCTATGTCCAACAGGAGCTTGATGATATCTCAAAATTG 3766
Qy      2291 GATCAAGAGAGCTTTTGTGATGAAGATGATCAAGCAGGAAAGTCTGGAATAATTCATT 2350
Db      3767 GCATGAGGAACTATTCAAGATGAGCCACTGATGGAGGAGACACAAAGAGGGAG 3826
Qy      2351 ATGATGATGCGGTATAGACAAATTCCTGA-----TCGTGATCTCGTGAGGCGAGG 2404
Db      3827 AAGATAGCAGTGTATCCACTAGCATGATAAGGCCATTGAACGCTGCTAGACCGTAACC 3886
Qy      2405 AAGTCTCAGTGGATGATGAGAGGAGATGATCTTTAAAGGCTTTCAGGTGCTTAATT 2464
Db      3887 AGATGAGACTGAAGACACAGAAATTCAGGCGCATGAATGAATATTGAGCTCAATCAAAG 3946
Qy      2465 TTGAATATATAGATGAAATGAGCAGCAGCAATTAGAGGCACAGAGAGTGCCTGCTGAA 2524
Db      3947 TGCCCGCATGATGTGTACGGGAAGAGAAATGCGGGAGGAGAGGATGAGACGGGAA 4006
Qy      2525 GCAAAATCTCAGCAGGCAATTCATGATAGACAAAGTTATTGCGGAAGAGTTGTTTAAAGATA 2584
Db      4007 TCATTAAACAGGAAGAAAGTGTGATCCTGA---CTACTGGGAGAAATTCCTGCGGCACC 4063
Qy      2585 AATTGAGCTGCACCGCTGAGGAGCTTAATGCTCTTGGAAGAAAGGAGAGAGTGCCTGCA 2644
Db      4064 ATTATGAGCAGCAGCAAGAAAGATCTAGCCCGAAATCTGGGCAAGGAAAGAAAGATCCGTA 4123
Qy      2645 AGCAGTTGGTATTCATTGAAG---AAGATGATCTTGTGCTGTTTGGAAAGATGTGAGCTCTG 2701
Db      4124 AACAGTCAACTACATGATGCTGCCAGAGGACCGAGATTGCGCAGGACACCATCGC 4183
Qy      2702 ATGAGATGAAGATTATGAGCTGAGTCAACAGATGTTGAAGCAGCAGGACAGAGATTC 2761
Db      4184 ACRAACAGTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTTGATGAACGTT 4243
Qy      2762 AGCGGTCGACGGCCGTCAGAGAAAGGTCGCGATATATTGGAACCAA----- 2812
Db      4244 CAGAAGCTCCCGTAGCCCGTAGTGGTGAAGGCGCTGCGGAATGATAAGAACCAATTGC 4303
Qy      2813 CTCGTTGATGGAAGGTGAGGGAGATCTTTTACAGATGACTGGGTTTCAACACAGAGTCAAA 2872
Db      4304 CTCCTCTGTGGCCCGTGTGGTGGGAATATTGAGTACTTGTGTTTAACTGCTCGTCAGC 4363
Qy      2873 GGGCCATTTTGTACAGACTTTGATGAGGTATG-----AGCTGGCAATTTTGATT 2923
Db      4364 GAAAGGCTTTCTTAATGCAATATTGCGGTATGATGCTATGCGACCTCAGGATGCTTTACTA 4423
Qy      2924 GGAAGGAGTTTGTTCCTCGCTTAAAGCAGAGAGACCTTTGAAGAAATAAATGAATATGGA 2983

```

```

Db      4424 CCAGTGGCTTGAAGACACCTCGCAGGCAATCAGAGAAAGAGTTCAAGGCATATGCT 4483
Qy      2984 TACTCTTCTTGAAGCAGCATGCTGAAGAAATAGACAGAAATTTCTCAACCTTTTCAGATG 3043
Db      4484 CTCTTTTCATGCGGCTTATGTCGCGCGGCGCAGATGGGCTGAGACCTTTGCTGATG 4543
Qy      3044 GTGTCGCCAAGAGAGACTTAGAATAGAGAGTGTCTAGTCAAGATGCTCTTCTCTATAC 3103
Db      4544 GTGTCGCCGAGAGGCGCTGCTCGCAGCATGCTCTTACTAGAAATGTTGTTATGCTT 4603
Qy      3104 TAGTTTCAGAGAGAGTGAATTTGTAGAGATCATCCAGGG 3144
Db      4604 TCATTTCGCAAGAGGTTTCAGGAGTTTGAACATGTTAATGGG 4644

RESULT 8
A56817 LOCUS A56817 6327 bp DNA linear PAT 03-MAR-1998
DEFINITION Sequence 1 from Patent WO9628547.
ACCESSION A56817
VERSION A56817.1 GI:3712830
KEYWORDS
SOURCE
ORGANISM
unidentified
unclassified.
REFERENCE
1 Seelig,H.P. and Renz,M.
AUTHORS DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
TITLE Patent: WO 9628547-A 1 19-SEP-1996;
JOURNAL PRIVATES INST FUER IMMUNOLOGIE (DE)
COMMENT Other publication DE 19509279 960515.
FEATURES
source
location/Qualifiers
1..6327
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
41..5738
/note="unnamed protein product"
/codon_start=3
/protein_id="CAA03424.1"
/db_xref="GI:3712831"
/db_xref="REMBL:CAA03424"
/translation="ASGLSPSPCSAGSEEDMDALLNNSLPPHPHNEEDPEEDLSE
TETPKLKKKKPKPRDKPIPKSKQKKRMILLCQLGSSGEGPFVEEEEAVALRS
DSEGDYTPGKKKKKLGPKKKSKRKEEDDDDDSKPKSKSQAQLLDDWNE
DIDHVFSEEDYRLTNYAFSQFVRPLIAAKPKIAVSKMMVLAGKWFSTNNPFK
GSSGASVAAAAAVALVESMTATEVAPPPEVPIRKAKTKEGKPNARRPKGS
PRVPAKPKPKKVAPLKILGFGSKRKRSSDDDDLOVESDFFDASINSVSDGS
TSRSRSRKLATTTKKKKEBEVTAQVETDHDYCEVCOGGRIILCDTCPRVH
MYCLDFDMEKPEGKWCSPHCBEKIQWEAKEDNSEGEILEEVGGDLSEEDHHEF
CRVCKDGGELCCDTCPSYHICLNPLPEIPNGEWLCPRCTCPALKKQVKLILWK
WQPPSPTPVPRPDADNTPSPKLEGRPERQFFVWQMGSMYHSCSWSELCLELHC
QVMFNRYORKNMDMBPPSGDFGDEBKSRKNKDPKPAEMBEERFYRGT KLPWMNII
RLNHSVDKKGHVHLYIKWRDLPYQASWESDEIQDYDLFKOSVYHRELNRGEG
RFGKLLKVKLRLKLERPPTPTVTVKYEQPEVLDATGTLHPQWMLGNLWRSW
AGQDTILADEMLGKTVQTAFLVLSLYKEGSKGFLVSLAPLSTLINWEREMWAP
DMYVTVGDKDSRAIIRNEHFSFDNAIRGGKSKMKKEASVKEHVLTSYELLFTI
DMAIYSIDWACLIVDEARLKNQSKFRLVNGSLQHKLLLTGTLQNNKPSKTELI
NLFPLPERHLEGGFEEPADIAKEDQIKLHMLGPHMLRLKADVPKNMPSKTELI
VVELSPMKQKVKYKILTRNFALNARGGNOVSLNVLNVMDLKCKCNHPVLPVAAVE
APKPMNGWDGSAIRASCKLLLOKMLKXEGGHEVLI FQMTKMLDLEDFLEHE
GKYRIDGITGNMRKEAIDRNPAGAOQFCFLSTRAGGLGINLATDVTIIVDS
WPHNDIQAFSAHRHIGONKVMYRFTVTRASVEERTIQVAKKMMMLTHLVVRPLGGS
ETEDTELQMNELSLFSGVAQVVRREEEGBEEVEREIIKQESVDYVWEKLLRHH
YEQOQEDLARNLKGKRIKQVYNDGSDODDQSDNODSDYVASEGDEDDPDE
RSEAPRPSRKLGRNDKDKPLPILLARVGNLEVLGFNARQKAFNAIMRWGMPQD
ATTQMLVRLDKGKSEKFAIVSLFMRLHCEPGADGATPADGVPREGLSRQHVLTIR
IGVMSLIRKQVQFEHVRGWSMELAEVEENKMSQSPGSPKTPPTPGTQNT
PAPVPAEDGIIKEENSLAKEIEGEKSTAPETAIECTQAPASEDEKVVVEP
PSGEKVEKAVERKEETPEPKGAADVEKEKSAIDLTPIVVEDKEEKKESEK
KEVMLQNGTPKLDNDEKOKNIKORFMENIADGGFTLHSLWNEERATVTKTYE
IWHRRHDYELLAGINHGVARQDIQNDPRYAILNEPFFKGEWNRGNFLEIKNFLARR

```

FKLLEQALVIEEQLRRAYALNMSBPSHPMALNTRFAEVECLAEHQHLSKESMAGN
KPNVILHVKLQLELLSDMKADVRLPATIARIIPVAVRLQWSEIRNLSRLANRAP
EPTPQQVAQQ"

ORIGIN

Query Match 10.1%; Score 421.4; DB 6; Length 6327;
Best Local Similarity 51.9%; Pred. No. 4.4e-88;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;

Qy	818	TACATCATACGAGCTTGAGGACTTAATTTTTTGGCGGTCTCGTGTGTCAAAACAGAGCGC	877
Db	2176	TGCACCCCTATCAATAGGAGCGCTGAATTTGGTTCGCTCTCTCGGCTCAGGCACTG	2235
Qy	878	ATGTAATCTTGTGATGAATGGAGCTAGCAAGACAAATTCAAAGCATTTGCCCTTTAG	937
Db	2236	ACACCATCTTGGCTGATGAGATGGGCTTGGGAAATCTGTACAGACAGAGCTCTTCGT	2295
Qy	938	CTTCACCTTTTGGAGG-----AACCTCATTCGCGCAATTTGGTAATTTGCTCCTCATCGA	991
Db	2296	ATTCCCTTTTCAAGAGGGTCAATTCCAAAGGCCCTTCTCTAGTGAGCGCCCTCTTTCTA	2355
Qy	992	CTCTGCGTAAGTGGAGAGAGAGTTTGCACATGCGGCCCCACAGATGAAGCTGGTTATGT	1051
Db	2356	CCATCATCAACTGGGAGCGGAGTTTGAATGTGGGCTCCAGACATGTATGTCTAACT	2415
Qy	1052	ATTTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAACATGATTTTACTTATCGAAAG	1111
Db	2416	ATGTGGGTGACAGGACAGCGCTGCCATCATCCGAGAGATGATTTCTCTTTGAAAGCA	2475
Qy	1112	ATCAAAAAGAGATCAAGAAAAGAAATCTGACAAATTAAGTAGGAAAGCAAGCAAAAA	1171
Db	2476	ATGCCATTCGTGGTGGCAAGAGCCCTC-----CCGCATGAAGAAAGAGGCAT	2523
Qy	1172	GAATCAAGTTGATGTCTCTCATCATCGTATGAGATGATCAACCTAGATTCAGCAGTTC	1231
Db	2524	CTGTGAATTCATGTGCTGCTGATCCTATGATGAATTTGATCACCATTGATGCTGCTATT	2583
Qy	1232	TAAAAACCAATTAAGTGGGAGTGCAATGTTGTGATGAAGGTCATCGACTGAATAAAGG	1291
Db	2584	TGGGCTCTATTGATTTGGGCTGCTCATCTGTTGATGAAGCCCATCGGCTGAGAACATC	2643
Qy	1292	ATTCAAAGCTGTTCTCTTCAATGACACAGATATTCAAGTAACCAACCGTATTCTTGACAG	1351
Db	2644	AGTCTAAGTTCTTCGGGTATTGAATGTTACTCACTCCAGCACAAAGCTGTTGCTGACTG	2703
Qy	1352	GAACACCACTTCAGAACAACTTTGATGAACATTTTTCATGCTCATGCATTTCTTGATCGG	1411
Db	2704	GGACACCAATTAACAAAACAACTGGAAGAGTTGTTTCTATCTGCTCAACCTTTCTCACCCCG	2763
Qy	1412	GGAAAGTTTGGAGTTTGGAGAGTTCCAGGAGGAGTTCAAGATATTAAATCAAGAGGAGC	1471
Db	2764	AGAGTTTCCAAATTTGGAAGTTTGGAGGAGTTTGGAGGAGTTTGGACATTTGCCNAGAGGACC	2823
Qy	1472	AGATCTCAAGGTTGCACAAAATTTGGCTCCACATTTTGTCTCAGAGGAGTAAAAAGACG	1531
Db	2824	AGATAAAAAAATCGATGACATGCTGGGGCGCACATGTTGCGCGGCTCAAGCCGATG	2883
Qy	1532	TAATGAAGACATGCCCCCAAAAAGAGCTCATTTTGGTGTGTTGATCTGAGCAGTCTGC	1591
Db	2884	TGTTCAAGAACATGCCCTTCCAAGACAGAACTTAATTTGCTGTGGAGCTGAGCCCTATGC	2943
Qy	1592	AGAAAGATATTACAAAGCTATTTTCCCGCTAATTTCAAGTATTGA-----CAAAAAAGG	1648
Db	2944	AGAAGAAATCTACAGTACATCTCTCACTCGAAATTTTGAAGCACTCAATGCCCGAGGTG	3003
Qy	1649	GAGGTGCTCAAAATTTCCCTTAATCAATATTATGATGGAATTAACGAAAAGTATGCTGCCATC	1708
Db	3004	GTGGCAACAGGTTGTTCTCTGCTGAATGTGGTATGATGATCTTTAAGAAAGTGTGCAACCATC	3063
Qy	1709	CTTA-----TATGCTAGAGGTTGTTGAGCCAGTTATTCACGACGCAATGAAGCTT	1759
Db	3064	CATACCTTCTCCCTGTGGCTCAATGGAAAGCTCCTAAGATGCCCTAATGGCATGTATGATG	3123

Qy	1760	TCAAAACAACCTTTTGGAGTCTTGTGAAAGCTGCAACTTCTAGATATAAAATGATGGTCAAAC	1819
Db	3124	GCAGTGCCCTAATCAGAGCATCTGGGAAATTTATGTGCTGCGAGAAAATGCTCAAGAACC	3183
Qy	1820	TGAAAGAGCAAGACACAGAGTCTTAATATACACAGTTTTCAGCATATGCTGAGACTTAC	1879
Db	3184	TTAAGGAGGTTGGCATCTGCTGCTACTCATCTTTTCCAGATGACCAAGATCTAGACCTGC	3243
Qy	1880	TTCAAGACTACTCTACCCATAAGAAATGGCAGTACGAGCGAATTTGATGGAAGGTTGGCG	1939
Db	3244	TAGAGGATTTCTTGGAACTGAAGGTTATAAATACGAACGCATCGATGGTGGAACTACTG	3303
Qy	1940	GAGCTCAGCGGCAAAATACGCATAGATCGGTTCAATGCGCAAAAATTTCTAAACAAGTTTGT	1999
Db	3304	GGAACTCGGCAAGAGGCCATTTGACCGCTTCAATGCACCGGTGCTCAGCAGTTCTGCT	3363
Qy	2000	TTTTGCTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA	2059
Db	3364	TCCTGCTTCCACTCGAGCTGGGGCTTTGGAATCAATCTGGCCACTGCTGACACAGTTA	3423
Qy	2060	TCATTTATGACAGTGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC	2119
Db	3424	TTATCTATGACTCTGACTGGAAACCCCATTAATGACATTTAGGCTTTAGCAGAGCTCAC	3483
Qy	2120	GACTTGGCCAAAACAAATAAGGTGATGATTTATAGGCTCTATAAACCGAGGACCATTTGAAG	2179
Db	3484	GGATTGGCAAAATAAAGGTAAATGATCTACCGGTTTGTGACCGCTGCGTCAGTGGAGG	3543
Qy	2180	AAAGGATGATGCAATTTGAATAAAGAAATGTTCTTAGAGCATCTTGTGTTGG-----	2234
Db	3544	AGCGCATACGCGAGTGGCAAGAAAGAAATGATGCTGACGCATCTAGTGTGCGGCTG	3603
Qy	2235	-----GAAACTCAAAACACAAAACATTAATCAGGAAGTTAGATGACATCATCAGGTATG	2290
Db	3604	GGCTGGGCTTCAAGACTGGATCTATGTCCAAACAGAGCTTGTATGATATCTCAATTTG	3663
Qy	2291	GATCAAGAGAGCTTTTTTGTAGTGAAGATGATGAAGCAGGAAAGTCTGGAATAATTCATT	2350
Db	3664	GCATGAGGAATTTCAAGGATGAAGCCATCTGATGGAGGAGGAGACAAACAAAGAGGAG	3723
Qy	2351	ATGATGATGGGCTATAGACAAATTTGCTTGA-----TCGTGATCTCGTGAGGACAGG	2404
Db	3724	AAGATGACAGTGTATCTCACTACGATGAAGCCATTAAGCGCTGCTAGACCGTAAAC	3783
Qy	2405	AGCTCTCAGTGGATGATGAGAGGAGAAATGCTTTTAAAGGCTTTTCAAGGTGGCTAAAT	2464
Db	3784	AGGATGAGCTGAAGACACAGAAATTCAGGGCATGAATGAATTTTGAAGCTCATCAAG	3843
Qy	2465	TTGAATATATAGATAAATGAGGACAGCATTTAGAGGCACAGAGAGTCTGCTGTAATA	2524
Db	3844	TGCCCCAGTATGTGTACGGGAAGAAATGGGGAGGAAGAGGAGGTAGAACGGGAAA	3903
Qy	2525	GCAATCTTCAGCAGGCAATTTCTGATAGCAAGTTATTTGGGAAGAGTTCTTAAAGATA	2584
Db	3904	TCATTAACAGGAAGAAGTGTGGATCTCTGA-----CTACTGGGAGAAATTTCTCGGCA	3960
Qy	2585	AATTTGAGCTGCCAGCGCTGAGAGCTTAATGCTCTTGAAAAAGGAAGAGAGTCCGCA	2644
Db	3961	ATTATGAGCAGCAGCAAGAAAGATCTAGCCCCGAAATCTGGGCAAGGAAGAAATCCGTA	4020
Qy	2645	AGCAGTTGGTATCCATTGAAG-----AAGATGATCTTGTGTTTGGAAAGATGTGAGCTCTG	2701
Db	4021	AACAGGTCAACTACAATGATGGCTCCCAGGAGGACCGAGATTTGACAGGACGACCGTCCG	4080
Qy	2702	ATGAGATGAAGTTATGAAGCTGAGTCAACAGATGATGAGCGNCGAGGACAGGAGTTTC	2761
Db	4081	ACAAACAGTCCGATTTACTCTAGTGGCTTTCAGAGGAAGGTGATGAAGACTTTTGTATGAAC	4140
Qy	2762	AGACGGTCCAGCGCCCTGACAGAAAGGCTCGCATATAATTTTGGAAACCAA-----	2812
Db	4141	CAGAGCTCCCGGTAGGCCCAAGTCTGTAAGGCGCTGCGGATGATGAAGATGAAGCATTCG	4200
Qy	2813	CTCCGTTGATGGAAGGTGAGGGAGATCTTTTCAGAGTACTGGGTTTCAACAGAGTCAAA	2872

Db 4201 CTCCTCTGTGGCCCGTGTGGGGAATATGAAGTACTTGTGTTAATGCTCGTCAGC 4260
QY 2873 GGGCCATTTTGTACAGACTTTTCATGAGGTATGG-----AGCTGGCAATTTTGATT 2923
Db 4261 GAAAAGCCTTCTTAATGCAATTAATGCGATATGATGCGCACCTCAGGATGCTTTACTA 4320
QY 2924 GGAAGAGTTTGTTCCTCGCTTAAAGCAGAAGACCTTTTGAAGAAATAAATGAATATGAA 2983
Db 4321 CCCAGTGGCTTGTAAAGAGACTCGAGGCAATCAGAGAAGAGTTCAGGCAATATGTCT 4380
QY 2984 TACTCTTCTGAAGCAGTTCGTGAAGAAATAGACGAGATTTCTCAACCTTTTCAGATG 3043
Db 4381 CTCTTTTCATGCGGCATTTATGTAGCGGGGCGAGATGGGCTGAGACCTTTTGCTGATG 4440
QY 3044 GTGTGCCCAAGGAAGGACTTTAGAATAGAAGATGTTCTAGTCAGAATTTGCTTCTGATAC 3103
Db 4441 GTGTCCCGGAGAGGCTGTCTGCCAGCATGCTCTTACTAGAAATTTGGTATGTATGCTT 4500
QY 3104 TAGTTCAGGAAGAGTGAATTTTGTAGAAGATCATCCAGGG 3144
Db 4501 TGATTCCGAAGAGGTTTCAGGAGTTTGAACATGTTAATGGG 4541

RESULT 9
LOCUS AR224019 6328 bp DNA linear PAT 26-SEP-2002
DEFINITION Sequence 1 from patent US 6440679.
ACCESSION AR224019
VERSION AR224019.1 GI:23332625
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 6328)
AUTHORS Seelig, H.P. and Renz, M.
TITLE Dermatomyositis-specific auto-antigen
JOURNAL Patent: US 6440679-A 1 27-AUG-2002;
FEATURES
source 1..6328
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 10.1%; Score 421.4; DB 6; Length 6328;
Best Local Similarity 51.9%; Pred. No. 4.4e-88;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;
QY 818 TACATCCATACCGCTTGAGGAGCTTAATTTTGGGGTCTCGTGGTCAAAAACAGAGC 877
Db 2177 TGCACCCCTATCAAAATGGAGGGCTGAATTTGGTTGCGCTTCTCCTGGGCTCAGGGCACTG 2236
QY 878 ATGTAATCCTTGCTGATGAATGGGACTAGGCAAGCAATTCAAAGCATTCGCCCTTTAG 937
Db 2237 ACACCATCTTGGCTGATGAGATGGGCTTGGGAAACTGTACAGACAGCAGTCTTCTGT 2296
QY 938 CTTCACTTTTGGAGG-----AACCTCATTCGCGATTTGGTAATTTGCTCCTCTATCGA 991
Db 2297 ATTCCCTTTACAGGAGGCTCATTCAAAGCCCTTCTTAGTACGCGCCCTTTCTA 2356
QY 992 CTCTGCGTAACTGGAGAGAGTTTGGCCATGCGGCCCCACAGATGAAGCTGGTTATGT 1051
Db 2357 CCATCATCAACTGGAGCGGAGTTTGAATGTGGGCTCCAGACATGATATGCTGAACCT 2416
QY 1052 ATTTGGCACTGCGCAGCTCGAGCAGTTATCAGAGACATGATTTTACTTATCGAAG 1111
Db 2417 ATGTGGGTGAACAAGCAGCGCTGCCATCATCGAGAGATGAGTTCTCTTTTGAAGACA 2476
QY 1112 ATCAAAAAAAGATCAAGAAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAGCAAAAA 1171
Db 2477 ATGCCATCTGGTGGCAAGAGGCTC-----CCGCATGAAGAAAGAGGCAT 2524
QY 1172 GAATCAAGTTTGTATGCTCTCTACATCGTATGAGATGATCAACCTAGATTTCAGAGTTC 1231

Db 2525 CTGTGAAATTCATGCTGCTGACATCTATGAATTAATGATACCATGATGATGCTATTT 2584
QY 1232 TAAAAACCAATTAAAGTGGGAGTGCATGATTGTTGATGAAGGTTCATCGACTGAAAAATAAGG 1291
Db 2585 TGGGCTCTATTGATTGGGCTGCTCATGCTGATGAAGCCCATCGCTGAAGAACAATC 2644
QY 1292 ATTCAAGCTGTTCTCTTCATGACACAGTATTAAGTAAACACCGTATTCCTTGTACAG 1351
Db 2645 AGTCTAAGTTCTTCGGGTATTGAATGGTTACTCACTCCAGCAACAAGCTGTTCCTGACTG 2704
QY 1352 GAACACCACTTCAGAAACAACCTTGGATGAACCTTTTCATGCTCATGCAATTTCTTGTATGCGG 1411
Db 2705 GACACCATTAACAACAATCTGGAAGAGTTGTTTCATCTGCTCAACTTCTCACCCCG 2764
QY 1412 GGAAGTTTGGAAAGTTTGGAGGAGTTCCAGGAGAGTTTCAAGATATTAATCAAGAGAGC 1471
Db 2765 AGAGGTTCCACAATTTGGAAGGTTTTCAGGAGAGTTTTCAGGAGAGTTTCCCAAGGAGACC 2824
QY 1472 AGATCTCAAGTTGTCACAAAATGTTGGCTCCACATTTGCTCAGAGGGTAAAAAAGAGC 1531
Db 2825 AGATAAAAAACTGTCATGATGCTGGGCGCGCACATGTTGCGCGGCTCAAAAGCCGATG 2884
QY 1532 TAATGAAGACATGCCCCCAAAAAGAGCTCAATTTTGGGTGTTGATCTGAGCAGTCTGC 1591
Db 2885 TGTTCAGAACATGCCCCCAAGACAGAACTAATTTGCTGTTGGAGCTGAGGCCCTATGC 2944
QY 1592 AGAAAAATATTACAAAGCTATTTTACCCGTAATTTATCAAGTATTGA---CAAAAAAGG 1648
Db 2945 AGAAGAAATACACAGTACATCTCCTCGAAATTTTGAAGCACTCAATGCCGAGGTG 3004
QY 1649 GAGGTCTCAAAATTTCCCTTAATAACATTAATGATGAAATTAACGAAAGATATGCTGCCATC 1708
Db 3005 GTGGCAACAGGTTCTCTGCTGAATGTGTGATGATCTTAAGAAGTGTCTGAACCATC 3064
QY 1709 CTTA-----TATGCTAGAGGGTTGAGCAGCTATTTCACGACGCAAAATGAAGCTT 1759
Db 3065 CATACCTCTTCCCTGTTGGCTGCAATGGAAGCTCTTAAGATGCCCTATGGCATGTATGATG 3124
QY 1760 TCAAAACAATTTTGGAGTCTTGTGAAAGTGCACCTCTAGATAAAATGATGGTCAAAC 1819
Db 3125 GCAGTGCCTTAATCAGAGCATCTGGGAAATTTATGCTGCTGCAGAAAAATGCTCAAGAACC 3184
QY 1820 TGAAGAGCAAGACACAGAGTCTTAATATACACAGTTTTCAGCATATGCTGAGACTTAC 1879
Db 3185 TTAAGAGGGTGGGCACTCGTACTCATCTTTTCCAGATGACCAAGATCTAGACCTGC 3244
QY 1880 TTGAAGACTACTGTACCCATAGAAATGGCAGTACGAGCAATTTGATGGAAGGTTTGGCG 1939
Db 3245 TAGAGATTTCCTTGGAACTGAAGGTTATAAATACGAACGATCGATGGTGGAACTACTG 3304
QY 1940 GAGCTGAGCGGCAAAATACGCATAGATCGGTTCATGCGCAAAAAATCTTAAACAAGTTTGT 1999
Db 3305 GGAACATGCGGCAAGAGCCATTTGACCGCTTCAATGACCGGGTGTCTCAGCAGTTCTGCT 3364
QY 2000 TTTTGTCTCCACAAGAGCTGTGGCTTAGGAATAAATTTGCAACGGCTGATACAGTAA 2059
Db 3365 TCTTGTCTTCCACTCGAGCTGGGGGCTTGGAAATCAATCTGGCCACTGTCTGACACAGTTA 3424
QY 2060 TCATTTATGACAGTGAATGGAATCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
Db 3425 TTATCTATGACTCTGACTGAAACCCCATTAATGACATTCAGGCTTTAGCAGAGCTCACC 3484
QY 2120 GACTTGGCCAAAACAAATAAGTGTATGATTTATAGGCTCATAAACCGAGGCCATTTGAAG 2179
Db 3485 GGATTTGGCAAAATAAAGAGTAATGATCTACCGGTTTGTGACCCGCTGCTAGTGGAG 3544
QY 2180 AAAGGATGATGAATTTGCTAAAAAAGAAATGTTCTAGAGCATCTTTGTTGG----- 2234
Db 3545 AGGCTATCAGCAGGTGGCAAGAAAGAAATGATGCTGACGCACTAGTGTGGCGGCTG 3604
QY 2235 -----GAAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGATCATCAGGTATG 2290

Db 3605 GGCTGGGCTCAAGACTGGATCTATGTCCAAACAGGAGCTTGATGATATCTCAAATTTG 3664
Qy 2291 GATCAAGAGGAGCTTTTCTAGTGAAGATGATGAAGCAGGAAAGCTCTGGAAAATTTCAATT 2350
Db 3665 GCATCGAGGAAGTATCAAGATGAAGCCACTGATGAGGAGGAGACAACAAGAGGGAG 3724
Qy 2351 ATGATGAGGGCTTATAGACAAATTTGCTGA-----TCGTGATCTCGTGAGGACGAGG 2404
Db 3725 AAGATGACAGTGTATCCACTACGATGAAGGCCATTGAAGCGCTCTAGACCGTAACC 3784
Qy 2405 AAGTCTCAGTGGATGATGAAGAGAGAAATGATTTCTTAAAGGCTTTCAAGGTGCTTAATT 2464
Db 3785 AGGATGAGACTGAAGACACAGAAATTCAGGCGCATGAATATTTGAGCTCAATTCAAAG 3844
Qy 2465 TTGAATATATAGTGAATAGGACGACAGCAATTTAGAGGACACAGAGAGTCTGCTGAAA 2524
Db 3845 TGGCCCCAGTATGTGGTACGGGAAGAAGAAATGGGGGAGGAGAGGATGAAGCGGAAA 3904
Qy 2525 GCAAATCTTCAGCAGGCAATTTCTGATAGCAAGTTATTTGGGAAGAGTTGTTAAAGATA 2584
Db 3905 TCATTAACAGGAAGAAGTGTGATCCTGA---CTACTGGAGAAATTCGTCGGGCACC 3961
Qy 2585 AATTGAGCTGCACACGCTGAGAGCTTAATGCTCTTTGAAAGAAAGAGAGAGTGCACA 2644
Db 3962 ATTATGAGCAGCAAGAAAGATCTAGCCGAAATCTGGGCAGAAAGAAAGAAATCCGTA 4021
Qy 2645 AGCAGTTGGTATCCATTGAAG---AAGATGATCTTGCTGTTGGAGAGTGTGAGCTCTG 2701
Db 4022 AACAGGTCNACTCAATGATGGCTCCAGGAGGACCAGAGATTTGGCAGGACGACCGTCCG 4081
Qy 2702 ATGAGAGTGAAGTTATGAAGCTGAGTCAACAGAGTGGTGAAGCAGCAGGAGAGTTC 2761
Db 4082 ACACACAGTCCGATTACTCAGTGGCTTCAGAGGAAGTGTGATGAAGCTTTGATGAAGCTT 4141
Qy 2762 AGCGGTCGACGCCGTCAGAGAAGAGGTCGCGATAATTTGGAACCAA-----2812
Db 4142 CAGAAAGCTCCCGTAGGCCAGTCTGAAGGCCCTGCGGAATGATAAAGATAAGCCATTGC 4201
Qy 2813 CTCCGTTGATGAAGTTCAGGGAGATCTTTCAGAGTACTGGGTTTCAACAGAGTCAAA 2872
Db 4202 CTCCTCTGTTGGCCGCTGTTGGTGGGAATATGAAGTACTTGGTTTAAATGCTCGTCAGC 4261
Qy 2873 GGGCCATTTTGTACAGACTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT 2923
Db 4262 GAARAGCTTTCTTAATGCAATTAATGATGATGATGTCACCCCTCAGGATGCTTTTACTA 4321
Qy 2924 GGAAGAGTGTGTTCTCGCTTAAAGCAGAGACCTTTGAGAAATAAATGAATATGGA 2983
Db 4322 CCCAGTGGCTTGTGAAGACCTCGGAGGCAATCAGAGAAGAGTTCGAAGGCATATGCT 4381
Qy 2984 TACTCTTCTGAAGCACAATGCTGAAGAAATAGACGAGAAATCTCCACCTTTTCAGATG 3043
Db 4382 CTCCTTTTCAATGCGCATTTATGTAGCGGGGCGAGATGGGGCTTGAGACCTTTTCTGATG 4441
Qy 3044 GTGTGCCCAAGAGAGGACTTAGAATAGAAATGTTCTAGTCAAGAAATGCTCTTCTGATAC 3103
Db 4442 GTGTCCCCGAGAGAGGCTGTCTCGCCAGCATGCTTACTAGAAATTTGGTGTATGCTT 4501
Qy 3104 TAGTTTCAAGAGAGGTTGAATTTGTAGAGATCATCCAGGG 3144
Db 4502 TGATTCGCAAGAGGTTTCAGGAGTTTGAACATGTTAATGGG 4542

RESULT 10
AR269336
LOCUS AR269336 6328 bp DNA linear PAT 10-APR-2003
DEFINITION Sequence 1 from patent US 650923.
ACCESSION AR269336
VERSION AR269336.1 GI:29700453
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.

REFERENCE 1 (bases 1 to 6328)
AUTHORS Seelig,H.P. and Renz,M.
TITLE Dermatomyositis-specific auto-antigen
JOURNAL Patent: US 650923-A 1 31-DEC-2002;
FEATURES Location/Qualifiers
source 1..6328
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 10.1%; Score 421.4; DB 6; Length 6328;
Best Local Similarity 51.9%; Pred. No. 4.4e-88;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;

Qy 818 TACATCCATACCACTTGAGGGAGCTTAATTTTTTGGGTTCTCTGTCGTCACAAACAGACGC 877
Db 2177 TGCACCCCTATCAATGGAGGGCTGAATTTGGTTGGCTTCTCTGGGCTCAGGCACTG 2236
Qy 878 ATGTAATCTTGTGATGAATGGGACTAGGCAAGCAATTTCAAAGCATTTGCCCTTTTAG 937
Db 2237 ACACCATCTTGTGATGATGGCTTTGGGAAACTGTACAGACAGCAGTCTTCTGT 2296
Qy 938 CTTCACTTTTGAAGAG-----AACCTCATTCGCCATTTGGTAATTTGCTCTCTATCGA 991
Db 2297 ATTCCCTTTTACAAGAGAGGGTCAATTCCAAAGGCCCTTCTCTAGTGAGCGCCCTTTCTA 2356
Qy 992 CTCTGGTAACTGGGAGAGAGTGTGGCCACATGGGCCCCACAGATGAACGTGTTATGT 1051
Db 2357 CCATCATCAACTGGGAGCGGAGTTTGAATGTGGGCTCCAGACATGATGTGTAACCT 2416
Qy 1052 ATTTTGGCACTGCGCAAGCTTCGAGCAGTTTATCAGAGAAACATAGTTTTTACTTATCGAAG 1111
Db 2417 ATGTGGGTGACAGGACAGCGGTCATCCGAGAGATGATTTCTCTTTGAAGACA 2476
Qy 1112 ATCAAAAAAGATCAAGAAAAAGAAATCTGGACAATAATAGTAGCAGAAAGCAAGCAAAAA 1171
Db 2477 ATGCCATTTCTGTGGCAAGAGGCCCTC-----CCGCATGAAGAAAGAGGCAT 2524
Qy 1172 GAATCAAGTTTGTATGTCCTCCACATCGTATCAGATGATCAACCTAGATTTCAGCAGTTC 1231
Db 2525 CTGTGAATTCATGTGCTGCTGACATCTTATGAATTTGATCACCATTGACATGGCTATTT 2584
Qy 1232 TAAAAACCAATTAAGTGGGAGTGCATGTTTGTGATGAAGGTCAATCGATGAAAAAATAAGG 1291
Db 2585 TGGGCTCTATTGATTGGGCTGCTCATCTGTGATGAAGCCCATCGGCTGAAGAAACAATC 2644
Qy 1292 ATTCAAGCTGTTCTCTTATTGACACAGTATTCAAGTAACCAACCGTATTCTTCTGACAG 1351
Db 2645 AGTCTAAGTTCTTCCGGTATTGAATGGTTACTCACTCCAGCACAAGCTGTTGCTGACTG 2704
Qy 1352 GAACACCACCTTCAGAACCACTTCGATGAACCTTTTTCATGCTCATGCAATTTTCTGATCGG 1411
Db 2705 GGAACCATTTACAAAAACAATCTGGAAGAGTTGTTTCACTGCTCAACTTTCTACCCCCG 2764
Qy 1412 GGAAGTTTGAAGTTTGGAGAGTTCCAGGAGGAGTTCAAAGATATTAATCAAGAGAGC 1471
Db 2765 AGAGTTTCCACAAATTTGGAAGGTTTTTGGAGAGTTTGTCTGACATTTGCCAAGGAGACC 2824
Qy 1472 AGATCTCAAGTTGCAAAAAATTTGGCTCCCATTTGCTCAGAGGTTAAAAAAGACG 1531
Db 2825 AGATAAAAAAAGTGCATGACATGTCGGGGCCGACATGTTTGGCGGCTCAAAAGCCGATG 2884
Qy 1532 TAATGAAGACATGCCCCCAAAAAAGGAGCTCAATTTTGGCGTTGATCTGAGCAGTCTGC 1591
Db 2885 TGTTCAGAAACATGCCCCCTCCAGACAGAACTAATTTGGCGTTGGAGCTGAGCCCTATGC 2944
Qy 1592 AGAAAGAAATTAACAAGCTATTTTACCCGTAATTTATCAAGTATTTGA---AAAAAAGG 1648
Db 2945 AGAAGAAATACTACAAGTACATCTCTCACTCGAAATTTTGAAGCACTCAATGCCCGAGGTG 3004
Qy 1649 GAGGTGCTCAAAATTTCCCTTAATAACATTTATGATGGAATTTACGAAAGATGTCGCCATC 1708
Db 3005 GTGGCAACCAAGGTGCTCTCTGCTGAATGTGGTGGATCTTAAAGAGTGTGCAACCATC 3064

Qy	2762	AGACGGGTGCACGGCGCTACAGAAAGAGGGTCCGCAATATTTTGGAAACCA	-----	2812
Db	4142	CAGAACTCCCGTAGGCCCAAGTCGTAAGGGCCCTGCGGAATGATAAAGATAAGCCATTGC	4201	
Qy	2813	CTCGTTGATGGAAGGTGAGGGAGATCTTTTCAGAGTACTGGGTTTCAACACGAGTCAAA	2872	
Db	4202	CTCTCTGTTGGCCGGTGTGGGAAATATCGAAGTACTTGGTTTAAATGCTCGTCAGC	4261	
Qy	2873	GGGCCATTTTGTACACAGACTTTGATGAGGTPATGG	-----AGCTGGCAATTTTGATT	2923
Db	4262	GA AAAAGCCCTTCTTAATGCAATTATGCGATATGCTATGCCACCTCAGGATGCTTTTACTA	4321	
Qy	2924	GGAGGAGTTTGTCTCCGCTTAAAGCAGAGACCTTTGAAGAAATAAATGAATATGGAA	2983	
Db	4322	CCCAGTGGCTTGTAAAGAGACCTCGAGGCAAAATCAGAGAAAGTAGTTCAAGGCATATGTCT	4381	
Qy	2984	TACTCTTTTGAAGCACAATTGCTGAAGAAATAGACAGAGAAATTCCTCAACCTTTTCAGATG	3043	
Db	4382	CTCTTTTCATGCGGCATTATGTGAGCCGGGGCAGATGGGCTGAGACCTTTTGTCTGATG	4441	
Qy	3044	GTGTGCCCAAGGAAGACTTAGAATAGAAGATTTCTTAGTCAGAAATGTCTTCTGTATAC	3103	
Db	4442	GTGTGCCCCAGAGAGGCCCTCTCTCGCAGCATGTCTCTTACTAGAAATGGTGTATGTCTT	4501	
Qy	3104	TAGTTCAGGAGAGAGGTCAAAATTTGTAGAGATCATCCAGG	3144	
Db	4502	TGATTCGACAGAAGGTTTCAGAGTTTGAACATGTTAATGGG	4542	
RESULT 11				
AR367436	AR367436	Sequence 1 from patent US 6329517.	DNA	linear
LOCUS	AR367436	6328 bp	DNA	linear
DEFINITION	AR367436	Sequence 1 from patent US 6329517.	DNA	linear
ACCESSION	AR367436	GI:34600634		
VERSION	AR367436.1	GI:34600634		
KEYWORDS	Unknown.			
SOURCE	Unknown.			
ORGANISM	Unclassified.			
REFERENCE	1 (bases 1 to 6328)			
AUTHORS	Seelig,H.P. and Renz,M.			
TITLE	Dermatomyositis-specific auto-antigen			
JOURNAL	Patent: US 6329517-A 1 11-DEC-2001;			
FEATURES	Location/Qualifiers			
source	1. 6328			
ORIGIN	/organism="unknown"			
	/mol_type="genomic DNA"			
Query Match	10.1%;	Score 421.4;	DB 6;	Length 6328;
Best Local Similarity	51.9%;	Pred. No. 4.4e-88;		
Matches 1236;	Conservative 0;	Mismatches 1076;	Indels 69;	Gaps 10;
Qy	818	TACATCATCAACAGCTTGAGGGACTTAAATTTTGGGGTTCTCGTGGTCAAAACAGACGC	877	
Db	2177	TGCACCCCTATCAATGGAGGGCTGAAATGGTGGCTTCTCTGGGCTCAGGCACTG	2236	
Qy	878	ATGTAATCCTTGCTGATGAAATGGGACTAGGCAAGACAATCAAGCAATTCGCCCTTTTAG	937	
Db	2237	ACACCATCTTGGCTGATGAGATGGGCTTGGGAAAACTGTACAGACAGCAGTCTTCTGT	2296	
Qy	938	CTTCACCTTTTGGAGGAG	-----AACTCATTCGCGATTTGGTAATTTGCTCTCTATCGA	991
Db	2297	ATTCCCTTTTACAGGAGGGTCATTCCAAAGGCCCTTCTTAGTAGGGCCCCCTTCTTCTA	2356	
Qy	992	CTCTGCGTAATCTGGGAGAGAGTTTGGCCACATGGGCCCCACAGATGAACGTGGTTATGT	1051	
Db	2357	CCATCATCAACTGGAGCGGAGTTTGAATGTGGGCTCCAGACATGATGTGCTGAACCT	2416	
Qy	1052	ATTTTGGCACTGGCGAAGCTCGACGAGTTATACAGAAATATGATGTTTACTTATCGAAG	1111	
Db	2417	ATGTGGGTGCAAGGACAGCGGTGCGCATCATCCGAGAGATGATGTTCTCTTTGGAAGACA	2476	

QY 1112 ATCAAAAAAGATCAAGAAAAAGAAATCTGGCAAAATAAGTAGCGAAAGCAAGCAAAAA 1171
|||
Db 2477 ATGCCATTCTGGTGGCAAGAGCCTC-----CCGCATGAAGAAAGAGCAT 2524
|||
QY 1172 GAATCAAGTTTGATGTCTCTCTACATCGTATGAGATGATCAACCTAGATTCAGCAGTTC 1231
|||
Db 2525 CTGTGAAATTCATGTCTGTGATGATCCCTATGAAATGATCACCATTGACATGCTGATTT 2584
|||
QY 1232 TAAAAACCAATTAAGTGGGAGTGCAATGTTGATGAAGGTTCATCGACTGAAAAAATAAGG 1291
|||
Db 2585 TGGGCTCTATTGATTTGGGCTGGCTCATCTGTTGATGAAGCCCATCGGCTGAAGAACATC 2644
|||
QY 1292 ATTCAAGCTGTCTCTCTCAATTTGACACAGTATTCAGATTAACCAACCGTATTCTTGACAG 1351
|||
Db 2645 AGTCTAAGTTCTTCGGGTATTGAATGTTTACTCACTCCAGCACAAAGCTGTGTGCTGACTG 2704
|||
QY 1352 GAACACACACTTCAGAACCAACTTTTGATGAACATTTTTCATGCTCATGCAATTTCTTGATCGGG 1411
|||
Db 2705 GGACACCATTTACAAAAAATCTGGAAGAGTTGTTTCTGCTCAACTTCTCACCCCGG 2764
|||
QY 1412 GGAAGTTTGGAAAGTTTGGAGAGTTCAGGAGGAGTTCAAAGATATTAAATCAAGAGGAGC 1471
|||
Db 2765 AGAGTTCCCAATTTGGAAGTTTGTGGAGGAGTTTGTGACATTTGCCAAGAGGACC 2824
|||
QY 1472 AGATCTCAAGGTTGCACAAAATGTTGGCTCCACATTTTCTCAGAAAGGTTAAAAAAGACG 1531
|||
Db 2825 AGATAAAAAAATCGATGACATGCTGGGGCGGCACATGTTGGCGGCTCAAGCCGATG 2884
|||
QY 1532 TAATGAAGACATGCCCGCCCAAAAGAGCTCATTTTGGCTGTTGATCTGAGCAGTCTGC 1591
|||
Db 2885 TGTTCAAGAACATGCCCTCAAGACAGAACTTAATTTGTCGTGTGGAGCTGAGCCCTATGC 2944
|||
QY 1592 AGAAAGAAATTTACAAAGCTATTATTTTACCGGTAAATTAATCAAGTATTCGA---CAAAAAGG 1648
|||
Db 2945 AGAAGAAATCTCAAGTACATCTCTCACTCAAAATTTTGAAGCACTCAATGCCCGAGGTG 3004
|||
QY 1649 GAGGTGCTCAAAATTTCCCTTAATAACATTTATGATGTAATTTACGAAATAGTATGCTGCCATC 1708
|||
Db 3005 GTGGCAACAGGTGTCTCTGTGAATGTTGGTGTGATGATCTTAAGAAAGTGTGCAACCATC 3064
|||
QY 1709 CTTA-----TATGCTAGAGGTTGTGAGCCAGTTTATTCAGACGCAATGAAGCTT 1759
|||
Db 3065 CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCTTAAGATGCCCTTAATGGCATGTATGATG 3124
|||
QY 1760 TCAAAACAATTTTGGAGTCTTGTGAAAAGCTGCAACTTCTAGATAAAATGATGTCAAAAC 1819
|||
Db 3125 GCAGTGCCTTAATCAGAGCATCTGGGAAATTTATGCTGTGCGAGAAAATGCTCAAGAAC 3184
|||
QY 1820 TGAAGAGCAGGACACAGAGTCTCTAATATACACAGATTTTCAGCATATGCTGGACTTAC 1879
|||
Db 3185 TTAAGGAGGTTGGCATCGTGTACTCATCTTTTCCAGATGACCAAGATGCTAGACCTGC 3244
|||
QY 1880 TTGAAGACTACTGTACCCATAAGAAATGGCGATGACAGCGAAATTTGATGGAAGTTGGCG 1939
|||
Db 3245 TAGAGGATTTCTTGGAACATGAAGGTTATAAATACGAACCGCATCGATGGTGGATCACTG 3304
|||
QY 1940 GAGCTGAGCGCAATATAGCATAGATCGGTTTCAATGCCAAAAATTTCTAAACAAGTTTGT 1999
|||
Db 3305 GGAACATGCGGCAAGAGCCATTGACCGTTTCAATGACCGGGTGTGACGAGTTTCTGCT 3364
|||
QY 2000 TTTTGTCTCCCAAGAGCTGGTGGCTTAGGAATAAATTTTGAACCGGCTGATACAGTAA 2059
|||
Db 3365 TCTTGTCTTCCACTCGAGCTGGGGCCTTGAATCAATCTGGCCACTGCTGACACAGTTA 3424
|||
QY 2060 TCATTATGACGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
|||
Db 3425 TTATCTATGACTCTGACTGGAACCCCAATAATGATCAATTCAGGCTTTTAGCAGAGCTCACC 3484
|||
QY 2120 GACTTGGCCAAACAAATAAGTGTATTTATAGGCTCATATAACCGAGGCCCATTTGAAG 2179
|||
Db 3485 GGATGGGCAAAATAAAAGGTAATGATCTACCGGTTTGTACCCCGTGTGCGTCAAGTGGAG 3544
|||
QY 2180 AAGGATGATGCAATTCACATAAAAGAAAAATGGTTCTAGAGCATCTTGTGTTGG----- 2234
|||

Db 3545 AGCGCATACGCGAGTGGCAAAAGAAAATGATGTGACGTCATAGTGTGGCGCTG 3604
|||
QY 2235 ----GAAACTCAAAACACAAAACATTAATCAGGAAGGTTAGATGACATCATCAGGTATG 2290
|||
Db 3605 GGCTGGGCTCAAGACTGGATCTATGTCCAACAGGAGCTTGTATGATATCTCTCAAAATTG 3664
|||
QY 2291 GATCAAGAGGAGCTTTTGTGTAGTGAAGATGATGAAGCAGGAAAGTCTGGAATAATTCAAT 2350
|||
Db 3665 GCACTGAGGAACTATTCAAGGATGAAGCCACTGATGGAGGAGGAGACACAAGAGGGAG 3724
|||
QY 2351 ATGATGATCGGCTATAGACAAATTCCTTGA-----TCGTGATCTCGTGGAGGAGAGG 2404
|||
Db 3725 AAGATAGCAGTGTATCTCACTACGATGAAGGCCATGAACGGCTGTAGACCGTAACC 3784
|||
QY 2405 AAGTCTCAGTGGATGAAGAGGAGATGGAATCTTTAAAGGCTTTCAAGGTGGCTAAAT 2464
|||
Db 3785 AGATGAGACTGAAGACACAGAAATTCGAGGCGATGAATGAATTTTGAAGCTCATCAAAG 3844
|||
QY 2465 TTCAATATATAGATGAAATGAGCCAGCAGCATTTAGAGGCACAGAGTCTGCTGCTGAAA 2524
|||
Db 3845 TGCCCAAGTATGTGTCGGAAGAAAGATGGGGAGGAGAGGAGGTAGAACGGGAAA 3904
|||
QY 2525 GCAAAATCTTACAGCGCAATTTCTGATAGACAAGTTATTGGGAAGAGTTGTTTAAAGATA 2584
|||
Db 3905 TCATTAAAACAGGAAGAAAGTGTGATCTCTGA---CTACTGGGAGAAATTCCTCGGCACC 3961
|||
QY 2585 AATTTGAGCTGACACAGGCTGAGGAGCTTAATCTCTTGGAAAAAGAGAGAGTCGCA 2644
|||
Db 3962 ATTATGAGCAGCAGCAGGAAGATCTAGCCGCAAAATCTGGGCAAGGAAAGAAATCCGTA 4021
|||
QY 2645 AGCAGTTGGTATCCATTTGAAG---AAGATGATCTTCTGTTTGGAGATGTGAGTCTG 2701
|||
Db 4022 AACAGGTCAACTACATGATGCTGCCAGGAGCAGAGATTGGCAGGACCAAGTCCG 4081
|||
QY 2702 ATGAGATGAAAGTTATGAAGTGAAGTCAACAGATGGTGAAGCAGCAGGACAAGAGTTC 2761
|||
Db 4082 ACAACAGTCCGATTTACTCAGTGGCTTCAGAGGAAGTGAAGAACTTTGATGAACGTT 4141
|||
QY 2762 AGACGGTGCAGCGCTACAGAAAGGTCGCGATATTTTGGAAACCAA----- 2812
|||
Db 4142 CAGAGCTCCCCGTAGCCCGCATGCTGAAGGCGCTCGGGAATGATAAGATAAGCCATTGC 4201
|||
QY 2813 CTCGTTTGAAGAGTGAAGGAGAGATCTTTCAGAGTACTGGGTTTCAACAGAGTCAAA 2872
|||
Db 4202 CTCTCTGTTGGCCGCTGTTGGGGAATTTGAAGTACTTGGTTTAACTGCTCGTCAGC 4261
|||
QY 2873 GGCCCATTTTGTACAGACTTTGATGAGGTATGG-----AGCTGGCAATTTTGAAT 2923
|||
Db 4262 GAAAAGCCTTCTTAATGCAATTTATGCGATATGCTATGCTGACCTCAGGATGCTTTACTA 4321
|||
QY 2924 GGAGGAGTTTGTCTCGCTTAAAGCAGAGAGACCTTTGAGAAATAAATGATATGAA 2983
|||
Db 4322 CCAGTGGCTTGAAGAGACCTCGAGGCAATCAGAGAAAGAGTTCAGGCAATGTCT 4381
|||
QY 2984 TACTCTTCTGAAGACATTTGCTGAAGAAATAGACGAGAATTTCTCAACCTTTTTCAGATG 3043
|||
Db 4382 CTCTTTTCATGCGCATTTATGTAGCCGGGCGAGATGGGCTGAGACCTTTGCTGATG 4441
|||
QY 3044 GTGTGCCCAAGGAAGCACTTTAGAAATAGAAAGATTTCTAGTCAAGATTTGCTCTGATAC 3103
|||
Db 4442 GTGTCCCGGAGAAAGGCTGTCTCGCCAGCATGCTCTTACTAGAAATTTGGTGTATGCTT 4501
|||
QY 3104 TAGTTCCAGGAGAGGTAATTTGTTAGAGATCATCCAGG 3144
|||
Db 4502 TGATTCGAAGAGGTTTCAGGAGTTTGAACATGTTAATGGG 4542
|||

RESULT 12
AX332238
LOCUS
DEFINITION
ACCESSION

AX332238 6417 bp DNA linear PAT 09-JAN-2002
Sequence 2747 from Patent WO0194629.
AX332238

```

VERSION      AX332238.1  GI:18122872
KEYWORDS
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1
AUTHORS      Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
              Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE        Cancer gene determination and therapeutic screening using signature
              gene sets
JOURNAL      Patent: WO 0194629-A 2747 13-DEC-2001;
              Avalon Pharmaceuticals (US)
FEATURES
  source     1. .6417
             /organism="Homo sapiens"
             /mol_type="unassigned DNA"
             /db_xref="taxon:9606"
ORIGIN
Query Match      10.1%; Score 421.4; DB 6; Length 6417;
Best Local Similarity 51.9%; Pred. No. 4.4e-88;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;
QY 818 TACATCCATACACAGCTGAGGAGCTTAATTTTTCGGGTTCTCTGCTGCTCAAAACAGACGC 877
DB 2266 TGCACCCCTATCAAAATGGAGGGCTGAATTTGGTTCGGCTTCTCTCGGCTCAGGGCACTG 2325
QY 878 ATGTAATCTCTGCTGATGAATGGGACTAGGCAAGACAATTCAAAGCAATTCGCCCTTTTAG 937
DB 2326 ACACCATCTGGCTGATGATGGGCTTGGGAAACTGTACAGACAGAGTCTTCTCTGT 2385
QY 938 CTTCACTTTTGGAGAG-----AACCTCAATCCGCATTTGGTAATTTGCTTCTCTATCGA 991
DB 2386 ATTCCCTTTACAAGAGGGGCTATTCCAAAGGCCCTTCCTAGTAGGCGCCCTCTTTCTA 2445
QY 992 CTCGCGTAACCTGGAGAGAGATTTGCCACATGGGCCCCACAGATGAAGCTGGTTATGT 1051
DB 2446 CCATCATCAACTGGGAGCGGGATTTGAAATGTGGGCTCCAGACATGTATGTCGTAACT 2505
QY 1052 ATTTTGGCACTGGCGAAGCTCGAGCAGTTATACAGAAACATGATGATTTTACTTATCGAAG 1111
DB 2506 ATGTGGGTGCAAGGACAGCGGTGCCATCATCCGAGAGATGATGTTCTCTTTGAAGACA 2565
QY 1112 ATCAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAGCAAAAAA 1171
DB 2566 ATGCCATTCGTGGTGGCAAGAGGCCCTC-----CCGCATGAAGAAAGAGGCAT 2613
QY 1172 GAATCAAGTTTGTATGTCCTCTCACATCGTATGAGATGATCAACCTAGATTCAGCAGTTC 1231
DB 2614 CTGTGAATTTCCATGTGCTGCTGACATCTCTATGAATTTGATCACCATTGACATGGCTATT 2673
QY 1232 TAAAAACCAATTAAGTGGAGTGCATGATTTGATGAAGTGCATCGACTGAAAAATAAGG 1291
DB 2674 TGGGCTCTATTGATTGGGCTGCTCATCGTGGATGAAGCCCATCGGCTGAAGACAATC 2733
QY 1292 ATTCAAAGCTGTTCTCTTCAATGACACAGTATTCAAGTAACCAACCGTATTCTTCTGACAG 1351
DB 2734 AGTCTAAGTTCTTCGGGTATTGAATGTTTACTCACTCCAGCACAAGCTGTGCTGACTG 2793
QY 1352 GAACACCACTTCAGACAACATCTGGATGAACCTTTTCAATGCTCATGCAATTTTCTTGATCGG 1411
DB 2794 GGACACCATTTACAAAAAATCTGGAAGAGTTGTTTCACTGCTCAACTTCTCACCCCGG 2853
QY 1412 GGAAGTTTGGAGTTTGGAGGAGTTCAGGAGGAGTTTCAAGATATTATCAAGAGAGGC 1471
DB 2854 AGAGTTTCCAAATTTGGAAGGTTTGTGGAGGAGTTTGTGACATTTGCCAAGAGAGACC 2913
QY 1472 AGATCTCAAGGTTTGCACAAATGTGGCTCCACATTTTGTCTCAGAAGGGTAAAAAAAGACG 1531
DB 2914 AGATAAAAAAAGTGCATGACATGCTGGGGCCGACATGTTTGGCGGCTCAAGCCGATG 2973
QY 1532 TAATGAAGACATGCCCCCCCCAAAAAGAGAGCTCAATTTTGGGTGTTGATCTGAGCAGTCTGC 1591
2974 TGTTCAGAACATGTCCTCCAGACAGAACTAATTTGTGCTGGAGCTGAGCCCTATGC 3033
1592 AGAAAGAAATATTACAAAGCTATTTTACCCGTAATTAATTATCAAGATTTGA----CAAAAAGG 1648
3034 AGAAGAAATATTACAAAGTACATCTCTCACTCGAAATTTTGAAGCACTCAATGCCGAGGTG 3093
1649 GAGGTGCTCAAAATTTCCCTTAATAACATTTATGATGAAATACGAAAGATGCTGCCATC 1708
3094 GTGGCAACAGGTGCTCTCTGCTGAATGTGTGATGATCTTAAAGAAATGCTGCAACCATC 3153
1709 CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTACAGCAGCAAAATGAAGCTT 1759
3154 CATACCTCTTCCTTGGCTGCAATGGAAGCTCTTAAGATGCCCTAATGGCATGTATGATG 3213
1760 TCAAAACAACCTTTTGGAGTCTTGTGAAAGCTGCAACTTCTAGATAAAATGATGGTCAAAAC 1819
3214 GCAGTGCCTTAATCAGAGCATCTGGGAAATTTATGCTGTCAGAGAAATGCTCAAGAAC 3273
1820 TGAAGAGCAAGGACACAGAGTCTTAATATACACACAGTTTCAAGATATGCTGGACTTAC 1879
3274 TTAAGAGGGTGGGCACTGCTACTCATCTTTTCCAGATGACCAAGATGCTAGACCTGC 3333
1880 TTGAAGACTACTGTACCCATAGAAATGCAGTACGAGCGAATTTGATGGAAGGTTGGCG 1939
3334 TAGAGATTTCTTGGAAACATGAAGTTATAATACGAACGATCGATGGTGGAAATCACTG 3393
1940 GAGCTCAGCGGCAAAATACGATAGATCGGTTCAAATGCCAAAAATTTCTAAAGTTTGT 1999
3394 GGAACATGGGCAAGAGGCCAATGACCGCTTCAATGACCGGTGCTCAGCAGTTCTGCT 3453
2000 TTTTGTCTCCACAGAGCTGGTGGCTTAGGAATAAATCTTGAACGGCTGATACAGTAA 2059
3454 TCTTGTCTTCCATCCAGCTGGGGCTTGGAAATCAATCTGGCACTGCTGACACAGTTA 3513
2060 TCATTTATCAGTACTGGAATCTCTGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
3514 TTATCTATGACTCTGACTGGAAACCCCAATTAAGCAATCAGGCTTTAGCAGAGCTCAC 3573
2120 GACTTGGCCAAACAAATAAGTGATGATTTATAGGCTCAATAACCGAGGACCAATTTGAAG 2179
3574 GGATTTGGCAAAATAAAGAGTAATGATCTACCGGTTTGTGACCGCTGGCTCAGTGGAG 3633
2180 AAAGATGATGCAATTTGCTAAAGAAATGTTCTTAGAGCATCTTGTGTTGG----- 2234
3634 AGCGCATCAGCAGGTGGCAAGAAAGAAATGATGCTGACGATCTAGTGTGGGCGCTG 3693
2235 ----GAACTCAAAACACAAACATTAATCAGGAAGATTTAGATGACATCATCAGTATG 2290
3694 GGCTGGGCTCCAGACTGGATCTATGTCCAAACAGGAGCTTTGATGATATCTCTCAAAATTG 3753
2291 GATCAAAAGGAGCTTTTGTAGTGAAGATGATGAAGCAGGAAAGTCTGGAATAATTCATT 2350
3754 GCACTCAGGAATTAATCAAGGATGAAGCCATGATGGAGGAGGAGACACAAGAGGGAG 3813
2351 ATGATGATCGGCTATPAGCAAAATTTGCTGA-----TCGTGATCTGTGGAGGAGAG 2404
3814 AAGATAGCAGTGTATCCACTACGATGAAGGCCCATTCGAAACCGCTGCTAGACCGTAACC 3873
2405 AAGTCTCAGTGTATGAAGAGGAGATGGATTCTTAAAGGCTTTCAAGGCTTTCAAGGTGGCTAATT 2464
3874 AGGATGAGACTGAAGACACAGAAATTCGAGGGGCAATGAATGAATTTTGAAGCTCAATCAAAG 3933
2465 TTGAATATATAGTAAAAATGAGGACAGCAGCATTTAGAGGACACAGAGAGTCGCTGCTGAAA 2524
3934 TGGCCCGATGATGTGTACCGGAAGAGAAATGGGGAGGAGAGGAGGTAGAACGGGAAA 3993
2525 GCAAAATCTTTCAGCAGGCAATCTGATPAGACGAAGTTATTGGGAAGAGTTGTTTAAAGATA 2584
3994 TCATTAACAGGAAGAAAGTGTGGATCTCTGA---CTACTGGGAGAAATTTGCTCGCGCAC 4050
2585 AATTTGAGCTGCACAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGAGAGAGTTCGCA 2644
```

Db 4051 ATTATGAGCAGCAAGAAGATCTAGCCGAAATCTGGCAAGAGAAAGAAATCCGTA 4110
Qy 2645 AGCAGTTGGTATCCATTGAAG---AAGATGATCTTGTCTGTTTGGAGAGATGTGAGCTCTG 2701
Db 4111 AACAGTCAACTACATGATGGCTCCAGGAGCAGGATGGCAGGACCAAGTCCG 4170
Qy 2702 ATGGAGATGAAGTTATGAAGCTGAGTCAACAGATGTTGAAGCAGCAGCAAGAGATTC 2761
Db 4171 ACAACCAAGTCCGATTAAGTCTAGTGGCTTCAGAGGAAGGTGATGAAGACTTTTGATGAACGTT 4230
Qy 2762 AGACGGTCAACGGCGCTACAGAGAAGAGGTCCGATATTTTGGACCAA----- 2812
Db 4231 CAGAAGCTCCCGCTAGCCAGTCTGAAGAGCCCTGCGGAATGATAAAGATAAGCCATTGC 4290
Qy 2813 CTCGGTTGATGAAGGTGAGGAGAGTCTTTACAGACTACTGGGTTTCAACACAGAGTCAAA 2872
Db 4291 CTCTCTGTTGGCCGCTGTTGGTGGGAATATGAGTACTTGGTTTAAATGCTCGTCAGC 4350
Qy 2873 GGCCATTGTTTACAGACTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT 2923
Db 4351 GAAAGCCTTCTTAATGCAATATGCGATATGATGTCACCTCAGGATGCTTTTACTA 4410
Qy 2924 GGAAGAGTTGTTCTCGCTTAAAGCAGAGACCTTTGAAGAAATGAATATGAA 2983
Db 4411 CCCAGTGGCTTGTAAAGAGACCTCGAGGCAATCAGAGAAGAGTTTCAAGGCATATGCT 4470
Qy 2984 TACTCTTTGAAGCAATGCTGAAGAAATAGACGAGAATCTTCAACCTTTTCAGATG 3043
Db 4471 CTCTTTTATCGGCATTTATGAGCGGGGCGAGATGGGCTGAGACCTTTTCTGATG 4530
Qy 3044 GTGTGCCCAAGGAGGACTTAGAATAGAAGATGTTCTAGTCAGAAATGCTCTTCTGATAC 3103
Db 4531 GTGTCCCCGAGAGAGGCTGTCTCGCCAGCATGTCTTACTAGAATTTGGTGTATGCTT 4590
Qy 3104 TAGTTAGGAGAGAGTGAATTTGTAGAATCATCCAGGG 3144
Db 4591 TGATTGCAAGAGAGGTTTCAAGAGTTTGAACATGTTAATGGG 4631

RESULT 13
HSM12218 HSM12218 6417 bp mRNA linear PRI 13-MAR-1996
LOCUS H. sapiens mRNA for 218kd Mi-2 protein.
DEFINITION X86691
ACCESSION X86691
VERSION X86691.1 GI:1107695
KEYWORDS helicase; Mi-2 gene; Mi-2 protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 6417)
AUTHORS Seelig,H.P., Moosbrugger,I., Ehrfeld,H., Fink,T., Renz,M. and Genth,E.
TITLE The major dermatomyositis-specific Mi-2 autoantigen is a presumed helicase involved in transcriptional activation

JOURNAL Arthritis Rheum. 38 (10), 1389-1399 (1995)
MEDLINE 96017437
PUBMED 7575689

REFERENCE 2 (bases 1 to 6417)
AUTHORS Seelig,H.P.
TITLE Direct Submission
JOURNAL Submitted (26-APR-1995) H.P. Seelig, Institute of Immunology & Molecular Genetics, Kriegsstrasse 99, D-76133 Karlsruhe, FRG

FEATURES
source
1. .6417
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="p13"
/cell_line="HeLa"
/clone_lib="lambda"
1. .6417

gene

CDS

/gene="Mi-2"
90..5828
/genes="Mi-2"
/codon_start=1
/product="Mi-2 protein"
/protein_id="CAA60384.1"
/db_xref="GI:1107696"
/db_xref="GOA:Q14839"
/db_xref="SWISS-PROT:Q14839"
/translation="MASGLGSPSCGASGEEDMDALLNLSLPPHPENBEDPEEDLS
ETETPKLKKKKPKPRDKPKSKKQKRMILLQGLDSSGEGPEFVEEREVALR
ESDIDHYTFEDYRTLTNKAQFQVRPLIAAKNPKIAVSKMMMLGAKREFSTNNPF
KSSGASVAAAAAVALVATEVAPPPPEVPIRKAKTKEGKGNAKPKPG
SPRPDAKPKKQVAPLKIKLGGFGSKRSSEDDDLDESDFDASINSVSDG
STSRSSRKRLRTKKKKKEEVEVAVDGYETHODYCEVCOQGEIILCDTCPRAY
HVLCDPMEKAPGKWSCHCEKEGIOMEAKEDNSEGEILLEEVGDLEERDHME
FCVKDGGELCCDTCSSYHILNLPPLPEIPNGEWLCPRTCTALKKVKQKILW
KWQPPSPFPVPRPPDADNTPSPKLEGRPERQFFVKGWMSYHSCSVSEIQLELH
CQMFENYQRKNDMDPPSGDFGDESKRKRKNKDPKFAEMEEERYGIKPEWMI
HRIILHSVKGKHVYLIKWRIPLDQASWESDEVIQDYDLFKQYVWHRLEMRGEE
GRGKKLKVKLRLKRLRPETPTVDPTVKYRQPEYLDATGTLHPYQWELNLFDS
MAQGTDTILADEMGLKVTQAVFLYSLYKEGSKGPFVLSAPLSLTINWEREFWA
PDMYVTVGDKDSRAIIRENEFSFEDNAIRGKKASRMKKEASVKFHLVLTSELT
IDMAILGSIDWACLIVDEARLKNQSKFVLNGYSLQHKLLLTQLQNNLEELFH
LNLFLTPEFHNLEGFLEFADIAKEDQIKLHMLGPHMLRKLKADVKNMPSKTEL
IVRVLSMPKKYKYVILTRNFALNARGQVSLNVMDLKKCNHPLYLFPVAM
EAPKPMGYDGSALIRASGKLLLOKMLNKEGHRVLI FSOYTKMLDLLDFLEH
EGYKVERIDGGITGNWQBAIDRFNAPGAQCFLLSTRAGGLGINLATDTVIIVDS
DWNPHNDIQAFSAHRIGQNKVMYIRFVIRASVEERIITQVAKKMLHLVVRPGLG
SKTSGMSKQELDDILKFGTEELFKDADTGGDNKEGESSVHYDDKATERLLDRHQ
DETDTELOQMEYLLSPKVAQYVREEMEEEEVEEIIKQESVDYDWBKLLRH
HYEQOEDLARNLKGKIRKOVYNDGQOEDMDQDSDNOSDYSAVEEGDEPFD
EBSERPRSRKLRNDKDKPLPLIARVGNIEVLGNAROKAFINAIMRGMPPQ
DAFTTQWLVRDLRGKSEKFKAYVSLPMRHLCEPGADGAETADGVPRGLSRQVLT
RIGVMSLIRKKVOEFHNGRWSMPELAEVKNKMSQPSGSPKPTPTSTGDTQPN
TPAPVPAEDGKIENSLEKEEISEGEKVSATPETALECTOAPAPESDKVVE
PPEGEKVEKAEVKETEPMETEPKGAADVEKEKSAIDLTPIVVEDEKEEKEE
KKEVMLONGETPKDLNDEKQKNIKORFMFNADGFTLHSLWNEERAATVTKTY
EJWHRHDYLLAGIINHGVARWODIQRVAILNEPPFKGMNRGNFLIKNKFLAR
RFKLQALVIEBQLRRAAYLNMSDEPSHPSMALNTRFAEVECLASHOHLKESVAG
NKPANAVLHKVLKLELLSDMKADVTRLPATIRIIPVAVRLQMSERNILSLANRA
PEPTPOQVAQQ"

ORIGIN

Query Match 10.1%; Score 421.4; DB 9; Length 6417;
Best Local Similarity 51.9%; Pred. No. 4.4e-88;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;
Qy 818 TACATCCATACCACTTGAGGAGCTTAATTTTTTGGGGTTCTCGTGGTCAAAACAGACGC 877
Db 2266 TGCACCCCTATCAATGGAGGCGCTGAATTTGGTTGCGCTTCTCTGGGCTCAGGCACTG 2325
Qy 878 ATGTATCTCTGTGATGAATGGAGCTAGGCAAGACAATCAAGCATTTGCCCTTTTAG 937
Db 2326 ACACCATCTTGGCTGATGAGATGGGCTTGGGAAAACGTACACAGACAGTCTTCTCGT 2385
Qy 938 CTTTCACTTTTGGAGGAG-----AACTCATTCGGCATTTGGTAATTTGCTCTCTATCGA 991
Db 2386 ATTCCCTTTTACAAGGAGGGTCATTTCAAAGGCCCTTCTCTAGTAGGCGCCCTCTTCTTA 2445
Qy 992 CTCTGCGTAACCTGGGAGAGAGTTTGGCCATCGATGGGCCCCACAGATGAAGCTGGTTATGT 1051
Db 2446 CCATCATCACTGGAGCGGAGTTTGAATGTGGGCTCCAGACATGATGTGCTGAACCT 2505
Qy 1052 ATTTTGGCACTGGCGAAGCTCGAGCAGTTATCAGAGAACATGATGTTTTTACTTATCGAAG 1111
Db 2506 ATGTGGGTGACAAGGACAGCGCTGCCATCATCCGAGAGAATGAGTTCTCTCTTTGAAGACA 2565
Qy 1112 ATCAAAAAAGATCAAGAAAAGAAATCTCGCAAAATAAGTAGCGAAAGCAAGCAAAAA 1171
Db 2566 ATGCCATTCGTGTGGTGGCAAGAGGCTC-----CCGCATGAAGAAAGAGCAT 2613

Qy	1172	GAATCAAGTTTGATGTCTCCTCATCATGTAATGAGATGATCAACCTTAGATTTCAGACAGTTC	1231
Db	2614	CTGTGAAATTCCATGTGTCTGTGATGATCACTCTATGTAATTAATGATCACCATTTGACATGGCTATTTT	2673
Qy	1232	TAAAACCAATTAAGTGTGGAGTGCATGATTTGTTGATGAAGTTCATCGACTGAAAAATAAGG	1291
Db	2674	TGGGCTCTAATTGATTGGGCTTGCCTCATCTGTGGATGAAGGCCATCTGGCTGAAGAACAATC	2733
Qy	1292	ATTCAAAAGCTGTCTCTTCATTTCAGACAGATATCCAAGTAAACCACCGTATTCTTCTGCACAG	1351
Db	2734	AGTCTAAGTTCTTCGGGTATTGAATGGTTACTCACTCCAGCAACAAGCTGTGTCTGACTG	2793
Qy	1352	GAAACACACTTCAGAAACAACTTTGATGAACAACTTTTCATGCTCATGCAATTTCTTGTATGCGG	1411
Db	2794	GGACACCAATTACAAAACAACTGGAAGAGTTGTGTTTCATCTGCTCAAATTTCTCACCCCGC	2853
Qy	1412	GGAGTTTTGGAAGTTTCGAGAGTTCCAGGAGGAGTTCAAAGATATTAATCAAGAGGAGC	1471
Db	2854	AGAGTTCCACAATTTGGAAGTTTTTTGGAGGAGTTTGTGTGACATTTGCCAAGGAGGACC	2913
Qy	1472	AGATCTCAAGGTTGCACAAAATGTGTGCTCCACAATTTTCTCAGAAAGGTTAAAAAAGACG	1531
Db	2914	AGATAAAAAAATGTCATGACATGCTGGGCGGCACATGTTGCGGCGCTCAAGCCGATG	2973
Qy	1532	TAATGAAAGACATGCCCCCCAAAAGAGAGCTCATTTTTGCGTGTGATCTGAGCAGTCTGC	1591
Db	2974	TGTTCAAGAACATGCCCTCCAAGACAGAACTAATTTGTGCGTGTGGAGCTGAGCCCTATGC	3033
Qy	1592	AGAAAGAAATTACAAAGCTATTTTTTACCCTGATTAATCAAGTATTGA---CAAAAAGG	1648
Db	3034	AGAAAGAAATACTACAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCCCAGGTTG	3093
Qy	1649	GAGGTGCTCAAAATTTCCCTTAATAACATTAATGATGGAAATTACGAAAAAGTATGCTGCCATC	1708
Db	3094	GTGCAACCAAGGTGCTCTGCTGTAATGTGGTATGGATCTTAAAGAGTGTGTCNAACCATC	3153
Qy	1709	CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTACGACGCGCAAAATGAAGCTTT	1759
Db	3154	CATACCTCTTCCCCTGTGGCTGCAATGGAAGCTCTCAAGATGGCTTAATGGCATGTATGATG	3213
Qy	1760	TCAAAACAATTTTGGAGTCTTTGGAAAGCTGCAACTTTTAGATAAAATGATGGTCAAAC	1819
Db	3214	GCAGTGCCTTAATCAGAGCATCTGGGAAATTTATTGCTGCTGCAGAAAATGCTCAAGAACC	3273
Qy	1820	TGAAAGACGAAGGACACAGAGTCTCTAATATACACACAGATTTTCAGCATATGCTGGACTTAC	1879
Db	3274	TTAAGHAGGGTGGGCATCTGTGTACTCATCTTTTCCAGATGACCAAGATGCTAGACCTGC	3333
Qy	1880	TTGAAGACTACTGTAPCCCATAGAANAATGGCAGTACGAGCGAAATTTGATGGAAAAGSTTGGCG	1939
Db	3334	TAGAGGATTTCTTGGAAACATGAAGTTTATAATACGAACGCATCGATGGTGGATCACTG	3393
Qy	1940	GAGCTGAGCGGCAAAATACGNATAGATCGGTTCAATGCGCAAAAATTTCTAAACAAGTTTGT	1999
Db	3394	GGAACTGCGGCAAGAGGCCATTTACCGCTTCAATGCAACCGGGTGCTCAGCAGTCTCTGCT	3453
Qy	2000	TTTTTGTCTCCACAAGAGCTGTGGCTTTAGGAATAAATCTTGGCAACGGCTGATACAGTAA	2059
Db	3454	TC TT TGTCTTTCAC TCTGAGCTGGGGGCTTGGAAATCAATCTGGCCACTGCTGACACAGTTA	3513
Qy	2060	TCATTTATGACAGTGACTGGAATCCTCATGTCTGATCTTCAAGCAATGGCTAGAGCTCATC	2119
Db	3514	TTATCTATGACTCTGACTGGAAACCCCATTAATGACATTTAGGCTTTTAGCAGAGCTCACC	3573
Qy	2120	GACTTGGCCAAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGGAGGCCAATTGAAG	2179
Db	3574	GGATTGGGCAAAAATAAAAAGGTAATGATCTACCGGTTTGTGACCCGTTGCGCTCAGTGGAGG	3633
Qy	2180	AAAGGATGATGCANTGACTAAAAGAAAATGGTTCTTAGAGCATCTTTGTTTGG-----	2234
Db	3634	AGCCGATCAGCAGGTTGGCAAGAAGAAAATGATGCTGACGCGATCTCTAGTGGTGGCGGCTG	3693
Qy	2235	----GAAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG	2290

Db	3694	GGCTGGGCTCCAAAGACTTGGATCTATGTCTCAAAACAGGAGCTTGTATGATATATCTCTCAAAATTTG	3753
Qy	2291	GATCAAAAGGAGCTTTTGTCTAGTGAAGCATGAAGCAGGAAAAGTCTGGAAAAAATTCATT	2350
Db	3754	GCACGTGAGGAACATATTCACGATGAGCCACTGATGAGGAGGAGAGACACAAAGAGGGAG	3813
Qy	2351	ATGATGATCGGGCTATAGACAAATTCCTTGA-----TCGTGATCTCGTGGAGGCAGAGG	2404
Db	3814	AAGATGACAGTGTATCCACTACGATGAAGGCCATTGAACGGCTGCTAGACCGTAACC	3873
Qy	2405	AAGTCTCAGTGGATGATGAAGAGGAGNATGGATCTTAAAGGCTTTCAAGGTCTTCAAGTGGCTAATT	2464
Db	3874	AGGATGAGACTGAAGACACAGAAATTCAGGGCATGAATGAATATTTGAGCTCATTTCAAG	3933
Qy	2465	TTGAATATATAGATGAAATGAGCAGCAGCATTTAGAGGCACAGAGAGTTCGCTGCTGAAA	2524
Db	3934	TGGCCAGTATGTGTTACGGGAAGAGAAATGGGGAGGAAGAGGAGTGAAGCGGAA	3993
Qy	2525	GCAAATCTTCAGCAGGCAATTCCTGATAGACAAAGTTATTGGGAAGAGTTGTTAAAAAGATA	2584
Db	3994	TCATTAAACAGGAAGAAAGTGTGGTCTGA---CTACTGGGAGAAATTCCTCGGCACC	4050
Qy	2585	AATTTGAGCTGCACAGGCTGAGAGGCTTAATGCTCTTGGAAAAAGAAAGAGAGTCCGA	2644
Db	4051	ATTATGAGCAGCAGACAGAAAGATCTAGCCCGAAATCTGGCCAAAGGAAAAAGAAATCCGTA	4110
Qy	2645	AGCAGTTGGTATCCATTGAAG---AAGATGATCTTCTGCTGGTTTGGAGAGTGTGAGCTCTG	2701
Db	4111	AACAGGTCAACTACAATGATGGTCCGAGAGGACCGAGATTTGGCAGGACGACCAGTCCG	4170
Qy	2702	ATGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAAGGAGTTC	2761
Db	4171	ACAACAGTCCGATTACTCAGTGCCTCAGAGGAAGTGTATGAAGACTTTTGATGAACGTT	4230
Qy	2762	AGCGGTGTCAGGCGCGTA CAGAAGAAAGGTCGCGANATTTTGGAAACCAA-----	2812
Db	4231	CAGAAAGTCCCGTAGGCCAGTCGTAAGGCGCTCGGGAATGATAAAGATAAGCCATTGC	4290
Qy	2813	CTCGTTGATGGAAGGTGAGGGAGATCTTTACAGGTACTCGGTTCACACGAGTCAAA	2872
Db	4291	CTCTCTGTTGGCCCGTGTGGTGGGAATATGAAGTACTTGGTTTTTAATGCTCGTCAGC	4350
Qy	2873	GGGCCATTTTGTACAGACTTTGATGAGGTATGG-----AGCTGGCAAATTTTGATT	2923
Db	4351	GAAAGCCTTCTTAATGCAATTTATGCGATATGGTATGCCACCTCAGGATGCTTTTACTA	4410
Qy	2924	GGAAGAGTTTGTCTCGCTTAAAGCAGAAGACCTTTTGAAGAAATAAATGAATATGGA	2983
Db	4411	CCCAAGTGGCTTGTAAAGAGACCTCGAGGCAAAATCAGAGAAAGAGTTCAGGACATATGCT	4470
Qy	2984	TACTCTTCTTGAAGCAGATTCCTGTAAGAAATAGACGAGNATTCCTCAACCTTTTCAGATG	3043
Db	4471	CTCTTTTCTATCGCGCATTTATGTAGCCGGGGGGAATGGGGCTGAGACCTTTGCTGATG	4530
Qy	3044	GTGTGCCAAAGGAGGACTTTAGAATAGAAGATGTTCTAGTCAGAATTCCTCTCTGATAC	3103
Db	4531	GTGTCCCCGAGAAAGGCTCTCTCGCAGCATGTCTTACTAGNATTTGGTGTATGTCTT	4590
Qy	3104	TAGTTTCAGGAGAAGGTGAAATTTGTAGAAGATCATCCAGG	3144
Db	4591	TGATTCGAAGAAGTTTCAGGAGTTTGAAACATGTTAATGGG	4631

RESULT 14

AR338834

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

ORGANISM

Unknown.

AR338834

Sequence 325 from patent US 6569662.

AR338834

AR338834.1

GI:33725691

Unknown.

6475 bp

DNA

linear

PAT 17-AUG-2003

Unclassified.									
REFERENCE	1 (bases 1 to 6475)								
AUTHORS	Tang, Y.T., Zhou, P., and Drmanac, R.T.								
TITLE	Nucleic acids and polypeptides								
JOURNAL	Patent: US 6569662-A 325 27-MAY-2003;								
FEATURES	Location/Qualifiers								
	1..6475								
	source								
/organism="unknown"									
/mol_type="genomic DNA"									
ORIGIN									
	Query Match	10.1%;	Score 421.4;	DB 6;	Length 6475;				
	Best Local Similarity	51.9%;	Pred. No. 4.4e-88;						
	Matches 1236;	Conservative	0;	Mismatches 1076;	Indels 69;	Gaps 10;			
Qy	818	TACATCCATACGAGCTGAGGAGCTTAATTTTTTGGCGTTCTCGTGTCAAAACAGACGC	877						
Db	2395	TGACCCCTATCAAAATGGAGGCGCTGAATTTGGTTGGCGCTTCTCTCGGCTCAGGGCACTG	2454						
Qy	878	ATGTAATCCTTGCTGATGAATGGACTAGGCAAGACAATTCAAAGCAATTCGCCCTTTAG	937						
Db	2455	ACACCATCTTGGCTGATGAGATGGCCCTTGGAAAACTGTACAGACAGAGCTTTCCTGT	2514						
Qy	938	CTTCACCTTTTGGAG- - - - - AACCTCATTCGCGCAATTTGGTAATTTGCTCTCTATCGA	991						
Db	2515	ATTCCCTTTACAGGAGGCTCATTTCCAAAGGCCCTTCTAGTGAGCGCCCTCTTTCTA	2574						
Qy	992	CTCTGCGTAATCGGAGAGAGATTTGCCATGTCAGTGGGCCCCACAGATGAAGCTGTATGT	1051						
Db	2575	CCATCATCAACTGGGAGCGGAGTTTGAATTTGGGCTCCAGACATATATGTCGTAACT	2634						
Qy	1052	ATTTTGGCACTGGCCAAAGCTGACAGTATACAGAAACATGATTTTACTTATCGAAAG	1111						
Db	2635	ATGTGGGTGACAGGACAGCGCTGCCATCATCCGAGAGATGATTTCTCTTTGAAGACA	2694						
Qy	1112	ATCAAAAAAGATCAAGAAAAAGAAATCTGGACAATAAGTAGCGAAAGCAAGCAAAAA	1171						
Db	2695	ATGCCATCTGGTGGCAAGAGGCCCTC- - - - - CCGCATGAAGAAAGAGGCAT	2742						
Qy	1172	GAATCAAGTTTGATGTCCTCTCATCTGATGATGATGATGATGATGATGATGATGATG	1231						
Db	2743	CTGTGAATTTCCATGTGCTGTCATCTTATGATGATGATGATGATGATGATGATGATG	2802						
Qy	1232	TAAACCAATTAAGTGGGAGTGCAATGTTGTGATGAAGGTTCATCGACTGAATAAAGG	1291						
Db	2803	TGGGCTCTATTTGATTTGGCTGCTCTCATCTGTTGATGAAGCCCATCGGCTGAAGAACATC	2862						
Qy	1292	ATTCAAAGCTGTTCTCTTATTTGACAGATTAATCAAGTAAACCAACCGTATTTCTTGACAG	1351						
Db	2863	AGTCTAAGTTCTTCGGGTATTGAATGTTTACTCACTCCAGCAACAGCTGTTGCTGACTG	2922						
Qy	1352	GAACACCACTTCAGAACAACTTGATGAAGAACTTTTTCATGCTCATGCAATTTCTTGATCGG	1411						
Db	2923	GGACACCAATTAACAAAAAATCTGGAAGAGTTGTTTCTATCTGCTCAACTTTCTACCCCGG	2982						
Qy	1412	GGAAGTTTGAAGTTTGGAGAGTTTCCAGGAGAGTTTCAAGATATTAATCAAGAGAGC	1471						
Db	2983	AGAGTTTCCAAATTTGAAGTTTGGAGAGTTTGGAGAGTTTGGAGAGTTTGGAGAGAGC	3042						
Qy	1472	AGATCTCAAGGTTGCACAAAATTTGCTCCACATTTGCTCAGAAAGGTTAAAAAAGACG	1531						
Db	3043	AGATAAAAAAATCGATGACATGCTGGGGCGGCACATGTTGCGCGCTCAAGCCGATG	3102						
Qy	1532	TAATGAAGACATGCCCCCAAAAGAGCTCATTTTGGCTGTTGATGCTGAGGAGTCTGC	1591						
Db	3103	TGTTCAAGAAATATGCCCTTCAAGACAGAACTTAATTTGCGGTGTGGAGCTGAGCCCTATGC	3162						
Qy	1592	AGAAAGAAATTTACAAAGCTATTTTACCCGTAATTAATCAAGTATGA- - - CAAAAAGG	1648						
Db	3163	AGAAAGAAATACAGTACATCTCTCACTCGAAATTTTGAAGCACTCAATGCCCGAGGTG	3222						
Qy	1649	GAGTGTCTCAAAATTTCCCTTAATTAACATTAATGATGGAATTAACGAAAAAGTATGTCGCATC	1708						

Db	3223	GTGGCAACACGAGTGTCTCTCTGCTGAATGTGTGTGATGATCTTAAGAAAGTGTCTGCAACCATC	3282						
Qy	1709	CTTA- - - - - TATGCTAGAGGGTGTGAGCGAGTTATTACAGCGCAAAATGAAGCTT	1759						
Db	3283	CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCTTAAGATGCCCTAATGGCATGTATGATG	3342						
Qy	1760	TCAACCAACTTTTGGAGTCTTGTGAAAAGCTGCAACTTCTTAGATAAAATGATGGTCAAAC	1819						
Db	3343	GCAATGCCCTAATCAGAGCATCTGGGAAATATTGTGCTGTCAGAAAAATGCTCAAGAAC	3402						
Qy	1820	TGAAAGAGAGGACACAGAGTCTTAATATACACAGTTTACAGTATATGCTGAGACTTAC	1879						
Db	3403	TTAAGGAGGTGGGCATCGGTACTCATCTTTTCCAGATGACCAAGATCTAGACCTGC	3462						
Qy	1880	TTCAAGACTACTCTACCCATAAGAAATGGCAGTACGAGCGAATTTGATGGAAAGGTTGGCG	1939						
Db	3463	TAGAGGATTTCTTGGAAACATGAAGTTTATAATACGAACGATCGATGTGGTAATCACTG	3522						
Qy	1940	GAGCTGAGCGGCAAAATACGCATAGATCGGTTCATGCTCAATGCCAAAAATTTCTAACAGTTTGT	1999						
Db	3523	GGAAATCGGCAAGAGGCCATTGACCGCTTCAATGACCGGGTGTCTACGAGTCTTGCT	3582						
Qy	2000	TTTTGTCTTCCACAGAGCTGTGGCTTAGGAATAAATCTTGCACCGCTGATACAGTAA	2059						
Db	3583	TCTTGCTTCCACTCGAGCTGGGGGCTTGGAAATCAATCTGGCCACTGTCTGACACAGTTA	3642						
Qy	2060	TCATTTATGACAGTGAATCCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC	2119						
Db	3643	TTATCTATGACTCTGACTGGAACCCCATTAATGACATTCAGGCTTTAGCAGAGCTCACC	3702						
Qy	2120	GACTTGGCCAAAACAAATAAGGTGATGATTTATAGGCTCTAATACCGAGGACCATTTGAAG	2179						
Db	3703	GGATTGGCAAAATAAAGGTAAATGATCTACCGGTTTGTGACCGCTGCGTCAAGTGGAGG	3762						
Qy	2180	AAAGGATGATCAATTTGACTAAAAAAAATGTTCTAGAGCATCTTGTGTTGG- - - - -	2234						
Db	3763	AGCGCATACGCGAGTGGCAAGAAATGATGCTGACGCATCTAGTGTGCGGCGCTG	3822						
Qy	2235	- - - - - GAACTCAAAACACAAAAATTAATCAGGAAGTTAGATGATCATCATCAGGTATG	2290						
Db	3823	GGCTGGGCTTCAAGACTGGATCTATGTCMAACAGAGCTTGAATGATATCTCTCAATTTG	3882						
Qy	2291	GATCAAGAGAGCTTTTGTGTAGTGAAGATGATGAAGCAGGAAAGTCTTGGAAAAATTCATT	2350						
Db	3883	GCACTGGAAGAACTATTCAAGGATGAAGCACTGATGAGGAGGAGACAACAAGAGAGGAG	3942						
Qy	2351	ATGATGATGCGGCTATAGACAAATTTGCTTGA- - - - - TCGTGTCTCTGAGGAGCAGAGG	2404						
Db	3943	AAGATAGCAGTGTATCCACTACGATGATAAGGCCATTGAACGGCTGCTAGACCGTAACC	4002						
Qy	2405	AAGTCTCAGTGGATGATGAAGAGGAGAAATGGAATTTCTTAAAGGCTTTTCAAGGTGGCTAAT	2464						
Db	4003	AGATGAGACTGAAGACACAGAAATTCAGGCGCATGAATGAATTTTGAAGCTCATTCAGAG	4062						
Qy	2465	TTGAAATATATAGATGAATAATGAGCAGCAGCATTTAGAGGCAACAGAGAGTCTGCTGAAA	2524						
Db	4063	TGCCCCAGTATGTGTACGGGAAGAAATGCGGGAGGAAGAGGAGTAGAAGCGGAAA	4122						
Qy	2525	GCAAACTTTCAGAGGCAATTTCTGATGAGCAAGTATTTCGGGAAGAGTTGTTTAAAGATA	2584						
Db	4123	TCATTAACAGGAAGAAAGTGTGGATCTGA- - - CTACTGGGAGAAATTTCTCGGCGACC	4179						
Qy	2585	AATTTGAGCTGCCACCGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGTGCACA	2644						
Db	4180	ATTATGAGCAGCAGCAAGAAAGATCTAGCCCCCAATCTGGGCAAGGAAGAAATCCGTA	4239						
Qy	2645	AGCAGTTGGTATCCATTGAAG- - - AAGATGATCTTGTGTTTGGAGAGTGTGAGCTCTG	2701						
Db	4240	AACAGGTCAACTACATGATGGCTCCAGGAGGACCGAGATTGGCAGGACGACAGTCCG	4299						
Qy	2702	ATGAGATGAAGTTTATGAGCTGATGATCAACAGATGTTGAAGCAGCAGGACAGGAGTTC	2761						
Db	4300	ACAACCAAGTCCGATTTACTAGTGGCTTACAGAGGAAGGTGATGAAGACTTTTGTGAACGTT	4359						

```
QY 2762 AGACGGGTGACGCGCGTACAGAAAGGGTGGCGATTAATTTGGAACCAA----- 2812
DB 4360 CAGAAGCTCCCGTAGGCCCGAGTCGTAAGGGCCCTGGCGAATGATAAAGATAAGCCATTGC 4419
QY 2813 CTCCTGTTGATGGAAGGTGAGGGGAGATCTTTCAGAGTACTTGGTTCACACAGAGTCAAA 2872
DB 4420 CTCCTGTTGGCCCGTGTGGTGGGAATTAAGAAGTACTTGGTTTTAATGCTGTCGACG 4479
QY 2873 GGGCCCAATTTGTACAGACTTTGTATGAGGTATGG-----AGCTGGCAATTTTGATT 2923
DB 4480 GAAAAGCCCTTTCTTAATGCAATATATCGATATGGTATGCCACTCAGGATGCTTTACTA 4539
QY 2924 GGAAGAGTGTGTTCTCGCTTAAGCAGAGACCTTTGAGAGAAATAAATGAATGGA 2983
DB 4540 CCAGTGGCTTGTAAAGAGACCTCGAGGCAAAATCAGAAAGAGTTCAGAGGCATATGCT 4599
QY 2984 TACTCTCTTGAAGCACATTTGCTGAAGAAATAGACGAGAATCTCCAACTTTTCAGATG 3043
DB 4600 CTCCTTTTCATGGCGCAATTTATGTAGCCGGGGCAGATGGGCTGAGACCTTTGCTGATG 4659
QY 3044 GTGTGCCCAAGGAAGGACTTTAGAATAAGAGATGTTCTAGTCAGAAATGCTCTTCTGATAC 3103
DB 4660 GTGTCCCGGAGAGGSCCTGTCTCGCAGCATGCTCTTACTAGAAATGGTGTATGCTT 4719
QY 3104 TAGTTCAGGAGAGGTGAATTTGTAGAGATCATCCAGGG 3144
DB 4720 TGATTCGCAAGAAGGTTTCAGGAGTTTGAACATGTTTAATGGG 4760

RESULT 15
AX924075 6331 bp DNA linear PAT 18-DEC-2003
LOCUS AX924075
DEFINITION Sequence 29 from Patent WO03080105.
ACCESSION AX924075
VERSION AX924075.1 GI:40217079
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Al-Mahmood,S., Colin,S. and Schneider,C.
TITLE Angiogenesis regulator genes, pharmaceutical preparations
containing same and uses thereof
JOURNAL Patent: WO 03080105-A 29 OCT-2003;
Gene Signal (PR)
FEATURES
source 1..6331
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 10.1%; Score 420.8; DB 6; Length 6331;
Best Local Similarity 52.3%; Pred. No. 6e-88;
Matches 1233; Conservative 0; Mismatches 1032; Indels 93; Gaps 10;

QY 810 AGGCTTTGTACATCCATACAGCTTCAGGAGCTTAATTTTTCGCGTTCCTCGTGGTCAAA 869
DB 2409 AGGCACTTCGACATGATCAGTTGAAGGCTGAATCTGCTACGCTTCTCTGGGCCCA 2468
QY 870 ACAGAGCATGTAATCTTCTGATGAAATGGGACTTAGGCAAGACAAATTCAAAGCATTCG 929
DB 2469 GGGCACTGACACCATTTCTAGCTGATGAGATGGGCTAGGCAAGACCATCAACCATCGT 2528
QY 930 CTTTGTAGCTTCACTTTTGTAGAGAACCTCA-----TTCCGCATTTGGTAAATTCCTCC 983
DB 2529 CTTCTCTTACTCACTCTACAAGAGGGGCCACACAAAAGGTCCCTTCTCTGTGAGTCCCC 2588
QY 984 TCTATCGACTCTCGTAACCTGGGAGAGAGATTTGACACATGGGCCCCACAGATGAACGT 1043
DB 2589 ACTCTTACCATCAATTAACCTGGGAGCGGAGTTCAGATGTGGGACCCCAAAATTTCTATGT 2648
```

```
QY 1044 GGTATATGATTTTGGCACTGCCCAAGCTCGACGAGTTATTCAGAGAACATGAGTTTACTTT 1103
DB 2649 GGTGACATACACGGGTGACAGGACAGCGCGGCATCATTCGTGGAATGAATTTCTCCTT 2708
QY 1104 ATCGAAGATCAAAAAGATCAAGAAAGAAATCTGGACAAATAAGTAGCGAAAGCAA 1163
DB 2709 TGAGGACAAATGCCATCAAGGGGGCAAGAAAGCTTTTAAGATGAAGAAG----- 2756
QY 1164 GCAAAAAAGAAATCAAGTTTGTATGTCCTCTCACATCGTATGAGATGATCAACCTAGATTTC 1223
DB 2757 GGAGGCACAGGTGAAGTTCCATGTTCTCTGACATCGTATGAGTATGATCACCATTGATCA 2816
QY 1224 AGCAGTTCTMAAACCAATTAAGTGGAGTGCATGATGTTGATGTAAGTGCATCGACTGAA 1283
DB 2817 GGCAGCACTTGGTTCATCTCGCTGGGCCCTGTCTTGTGGTAGATGAGGGCCATCGACTCAA 2876
QY 1284 AATAAGGATTCAAAGCTGTTCTCTTATTGACACAGATTAATCAAGTAAACCCAGTATCTT 1343
DB 2877 GAACAAACAGTCCAAGTTTTTTCAGGGTCTCAATGGTTACAAGATAGATCATAAAGTTGCT 2936
QY 1344 TCTGACAGGAACACCACTTCAGAACAACTTGGATGAACCTTTTCATGCTCATGCAATTTCT 1403
DB 2937 GCTGACAGGAACCCCATTTGCAGATAAATCTGGAGAGCTCTTCCATCTCTCTGAACITTCCT 2996
QY 1404 TGATCGGGGAAAGTTTGGAAAGTTTGGAGAGTTTCAGAGGAGTTTCAGAGGAGTTTCAAGATATTAATCA 1463
DB 2997 CACCCACAGAGAGATTTAAACAACCTTGGAGGGCTTCTCGAGGAGTTTGTGTGACATATCCAA 3056
QY 1464 AGAGGAGCAGATCTCAAGGTTGCACAAATGTTGGTCCACATTTGCTCAGAGGGTAA 1523
DB 3057 AGAGGACAGATCAAGAAACTGCATGATTTGCTGGGGCCACACATGCTCGGAGACTCAA 3116
QY 1524 AAAAGACGTAATGAAGACATGCCCCCAAAAGAGAGTCTATTTTCGCTGTTGATCTGAG 1583
DB 3117 GGCAGATGCTTTAAGAACATGCCAGCCAAAGACAGAGCTCATCGTTTCGGTGGAGCTAAG 3176
QY 1584 CAGTCTGAGAAAGAAATATTAACAAAGCTATTTTACCCTTAATTAATTAAGTATTTGA---C 1640
DB 3177 CCCCATGCAGAAAGAAATATTAACAAATATCATCTGACTCGAAATTTTGGAGCCCTTGAATTC 3236
QY 1641 AAAAAGGAGGCTGCTCAAAATTTCCCTTAATAACATTAATGATGGAATTAACGAAAGTATG 1700
DB 3237 AGGAGTGTGGGAACCAAGGTCGCTGCTTAATATCATGATGGATCTTAAGAAAGTGGTG 3296
QY 1701 CTGCCATCTTATATGCTAGAGGGTGTG-----AGCCAGTTTATTCAGACACAAA 1751
DB 3297 CAACCATCCATACCTTTTCCCGTGCCTGCTATGGAGTCCCCCAAACTCCCAAGTGGGC 3356
QY 1752 TGAAGCTTTCAAAACAACTTTTGGAGTCTTTGCGAAGCTGCAACTTCTAGATAAAATGAT 1811
DB 3357 TTATGAGGTTGGGGCACTTATTAAGTCGCTCGGAAAGCTCATGCTCTCCAGAAGATGCT 3416
QY 1812 GGTCAAACTGAAAGAGCAGGACACAGAGTCCCTAATATACACAGATTTTCAGCATATGCT 1871
DB 3417 GCGAAAGCTGAAAGGAGCAGGACACCGAGTGTCTATCTTCTCGAGATGACCAAAATGTT 3476
QY 1872 GGACTTACTTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGCAATTTGATGAAA 1931
DB 3477 AGACTTGTCTGAGGACTTCTTAGACTATGAAGGCTACAGATGATGAGCGCATCGATGGTG 3536
QY 1932 GGTTCGGAGCTGAGCGCAAAATACGCATAGATCGGTTCAATGCGCAAAATTTCTAACHA 1991
DB 3537 TATCAGCGGTGCCCTGAGGACAGGAGCCATCGATCGGTTTAATGCTCTCTGGGGCCCAACA 3596
QY 1992 GTTTTGTGTTTGTCTCTCCACAAGAGCTGGTGGCTTAGGAAATAAATCTTGCACCGCTCA 2051
DB 3597 ATTCTGCTTCTCTCTGTCACCCGAGCTGGGGCCCTGGGCATCAATCTGGCCACTGCTGA 3656
QY 2052 TACAGTAATCATTTATGACAGTGAATCCTCATGCTGATCTTCAAGCAATGGCTAG 2111
DB 3657 CACTGTCACTCTTTGATTTCTGACTGGNAACCCCAATATGACATCCAGGCCCTTTAGCCG 3716
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 11:18:36 ; Search time 1474 Seconds
(without alignments)
12038.483 Million cell updates/sec

Title: US-10-049-137-1
Perfect score: 4177
Sequence: 1 acgagtagttgttgagag.....caactggtaaatcaagattc 4177

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124093041 residues
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4177	100.0	4177	4	AAD02817 Arabidops
2	850.8	20.4	3204	6	ABZ14721 Arabidops
3	424.6	10.2	6327	2	AAT32301 Dermatomy
4	421.4	10.1	5739	6	ABK70271 Human lun
5	421.4	10.1	6417	6	ABL64410 Stomach c
6	421.4	10.1	6417	6	ABX04169 Human mRN
7	421.4	10.1	6417	6	ABK84696 Human chr
8	421.4	10.1	6417	7	AAD54631 Human chr
9	421.4	10.1	6417	7	AAD54632 Human chr
10	421.4	10.1	6475	8	Aai58444 Human pol
11	421.4	10.1	6475	8	ADB48415 Novel hum
12	420.8	10.1	6331	5	ABV23250 Human pro
13	420.8	10.1	6331	7	AAD54629 Human chr
14	420.8	10.1	6331	7	ACD13374 Human DNA
15	409.5	9.8	7805	7	ABX63370 Human chr
16	387	9.3	3716	9	ADB23222 Human chr
17	387	9.3	4653	6	ABQ99469 Human cod
18	387	9.3	7849	7	AAD54630 Human chr
19	352.6	8.4	6553	4	Aai60230 Human pol
20	324.8	7.8	2673	4	ABL29909 Drosophil
21	324.8	7.8	4705	4	ABL29908 Drosophil
22	311.6	7.5	7303	4	ABL29648 Drosophil
23	296.6	7.1	6608	2	AAT42751 Chicken C

ALIGNMENTS

RESULT 1

AAD02817	
ID	AAD02817 standard; cDNA; 4177 BP.
XX	
AC	AAD02817;
XX	
DT	31-MAY-2001 (first entry)
XX	
DE	Arabidopsis thaliana PKL cDNA.
XX	
KW	PKL; chromo domain; helicase domain; DNA binding domain; CHD3 protein;
KW	zinc finger domain; transgenic plant; chromatin remodelling factor; LEC1;
KW	pickle root; developmental identity; embryo development; ss.
XX	
OS	Arabidopsis thaliana.
XX	
FH	Key
FT	CDS
FT	1..4155
FT	/*tag= a
FT	/product= "Arabidopsis thaliana PKL protein"
FT	145..288
FT	/*tag= b
FT	/note= "Encodes zinc finger domain; this region is
FT	specifically referred in claim 5"
FT	misc_feature
FT	343..453
FT	/*tag= c
FT	/note= "Encodes chromo domain -I; this region is
FT	specifically referred in claim 4"
FT	misc_feature
FT	571..681
FT	/*tag= d
FT	/note= "Encodes chromo domain -II; this region is
FT	specifically referred in claim 6"
FT	misc_feature
FT	877..2217
FT	/*tag= e
FT	/note= "Encodes helicase domain; this region is
FT	specifically referred in claim 4"
FT	misc_feature
FT	3205..3285
FT	/*tag= f
FT	/note= "Encodes DNA binding domain; this region is
FT	specifically referred in claim 4"
XX	
PN	WO200114519-A2.
XX	
PD	01-MAR-2001.
XX	
PF	18-AUG-2000; 2000WO-US022725.

Aad54627 Human chr
Aad54626 Human chr
Aca04019 cDNA down
Aad54628 Human chr
Abl06443 Drosophil
Abl06442 Drosophil
Abl06515 Drosophil
Abl06514 Drosophil
Abq75913 CHD activ
Aal48805 Murine an
Acf36116 Human chr
Abz14466 Arabidops
Aah15817 Human chr
Abl88441 Pain regu
Abx70887 Novel hum
Abk65338 Arabidops
Add31010 Plant yie
Aad54638 Human chr
Aad54637 Human chr
Aad54638 Human chr
Aad54641 Human chr

XX	20-AUG-1999; 99US-0149975P.	QY	541	GATTTTGTGCCATACGTCCTGAGTGGAGCCACCTGTTGATCGGATTCCTTGCTCGAGAG	600
PR	(PURD) PURDUE RES FOUND.	DB	541	GATTTTGTGCCATACGTCCTGAGTGGAGCCACCTGTTGATCGGATTCCTTGCTCGAGAG	600
PA					
XX	Ogas JP, Somerville CR;	QY	601	GAAGATGGGGAGCTGGAATATCTTGTCAAAATATAAAGAGCTATCCCTATGATGAATGTTAT	660
XX	WPI; 2001-218433/22.	DB	601	GAAGATGGGGAGCTGGAATATCTTGTCAAAATATAAAGAGCTATCCCTATGATGAATGTTAT	660
DR	P-PSDB; AAY72755.				
XX	Regulating plant developmental identity comprises transforming a host	QY	661	TGGGAGCTCAGNATCAGACATCTCAACCTTCCAGAAATGAAATTCAAAGGTTCAAGGATGTA	720
PT	cell with a nucleic acid encoding a PKL protein having at least one	DB	661	TGGGAGCTCAGNATCAGACATCTCAACCTTCCAGAAATGAAATTCAAAGGTTCAAGGATGTA	720
PT	chromo domain, a helicase domain and a DNA binding domain.	QY	721	AATTCTAGAACTCGCAGAAAGTAAAGATGTTGACCATATAAAGAAATCCAGAGACTTTCAA	780
XX	Claim 55; Page 67-76; 87pp; English.	DB	721	AATTCTAGAACTCGCAGAAAGTAAAGATGTTGACCATATAAAGAAATCCAGAGACTTTCAA	780
XX	The invention relates to a method of regulating plant developmental	QY	781	CAGTTTGATCATACTCCTGAATTCCTCAAGGGCTGTTACATCATCATCAAGCTTTGAGGGA	840
CC	identity which involves transforming a host cell with a nucleic acid	DB	781	CAGTTTGATCATACTCCTGAATTCCTCAAGGGCTGTTACATCATCATCAAGCTTTGAGGGA	840
CC	encoding PKL (pickle) protein having at least one chromo domain, a				
CC	helicase domain, a DNA binding domain and a zinc finger domain. This	QY	841	CTTAATTTTTCGGGTTCTCGTGTCAAAACAGACGCATCTAATCTTCTGCTGATGAAATG	900
CC	method is useful for transforming a host cell with PKL proteins for	DB	841	CTTAATTTTTCGGGTTCTCGTGTCAAAACAGACGCATCTAATCTTCTGCTGATGAAATG	900
CC	regulating developmental identity, such as for regulating the transition	QY	901	GGACTAGGCAAGACAAATTCAAAGCATTCGCCCTTTTAGCTTCACTTTTTCAGGAGAACCTC	960
CC	from embryonic to post-embryonic development. The PKL proteins act as	DB	901	GGACTAGGCAAGACAAATTCAAAGCATTCGCCCTTTTAGCTTCACTTTTTCAGGAGAACCTC	960
CC	chromatin remodelling factors to repress transcription of LEC1, a protein				
CC	that plays a role in regulating embryo development. The method is also	QY	961	ATTCCGCAATTTGGTAAATGCTCCTCTATCGACTCTGCTAACTCGGAGAGAGATTTGCC	1020
CC	used for generating transgenic plants expressing PKL proteins. The	DB	961	ATTCCGCAATTTGGTAAATGCTCCTCTATCGACTCTGCTAACTCGGAGAGAGATTTGCC	1020
CC	present sequence is a cDNA coding for Arabidopsis thaliana PKL protein	QY	1021	ACATGGGCCCCACAGATGAACGTTGTTATGTTTGGCACTCGCAAGCTCCGACGAGTT	1080
CC	found in pickle root. PKL functions in regulating developmental identity	DB	1021	ACATGGGCCCCACAGATGAACGTTGTTATGTTTGGCACTCGCAAGCTCCGACGAGTT	1080
CC	in host cells. It is classified as CHD3 protein based on the presence of	QY	1081	ATCAGAGAACATCAGTCTTACTTATCGAAGATCAAAAAAGATCAAGAAAAAGAAATCT	1140
CC	chromo domain, SNF2-related helicase/AtPase domain, DNA-binding domain	DB	1081	ATCAGAGAACATCAGTCTTACTTATCGAAGATCAAAAAAGATCAAGAAAAAGAAATCT	1140
CC	and PHD zinc finger domain				
XX					
SQ	Sequence 4177 BP; 1353 A; 729 C; 1018 G; 1077 T; 0 U; 0 Other;	QY	1141	GGCAAAATAGTAGCGAAACGAAAGCAAAAAGAAATCAAGTTTGATGTCCTCCTCACATCG	1200
		DB	1141	GGCAAAATAGTAGCGAAACGAAAGCAAAAAGAAATCAAGTTTGATGTCCTCCTCACATCG	1200
		QY	1201	TATGAGATGATCAACCTAGATTCCAGAGTTCTAAAAACCAATTAAGTGGGAGTGATGATT	1260
		DB	1201	TATGAGATGATCAACCTAGATTCCAGAGTTCTAAAAACCAATTAAGTGGGAGTGATGATT	1260
		QY	1261	GTTGATGAAGGTCATCGACTGAAAAATTAAGGATTCAAAGCTGTTCTTCTTCAATTGACACAG	1320
		DB	1261	GTTGATGAAGGTCATCGACTGAAAAATTAAGGATTCAAAGCTGTTCTTCTTCAATTGACACAG	1320
		QY	1321	TATTCAGATCAACCCGATTCCTTCGACGGAACACCCTTCAGAAACAACCTTGGATGAA	1380
		DB	1321	TATTCAGATCAACCCGATTCCTTCGACGGAACACCCTTCAGAAACAACCTTGGATGAA	1380
		QY	1381	CTTTTCATGCTCATGCAATTTCTTGATCGGGGAAGTTTGGAGTTTGGAGGAGTCCAG	1440
		DB	1381	CTTTTCATGCTCATGCAATTTCTTGATCGGGGAAGTTTGGAGTTTGGAGGAGTCCAG	1440
		QY	1441	GAGGAGTTCAAAGATATTAATCAAGAGGAGCAGATCTCAAGGTTGCAAAAAATGTTGGCT	1500
		DB	1441	GAGGAGTTCAAAGATATTAATCAAGAGGAGCAGATCTCAAGGTTGCAAAAAATGTTGGCT	1500
		QY	1501	CCACATTTGCTCAGAGGGTAAAAAAGAGCGTAATGAAGACATGCCCCCAAAAGGAG	1560
		DB	1501	CCACATTTGCTCAGAGGGTAAAAAAGAGCGTAATGAAGACATGCCCCCAAAAGGAG	1560
		QY	1561	CTCATTTTGGCTGTTGATCTGAGCAGTCTGACAGAAAGATATTACAAAGCTATTTTACC	1620
		DB	1561	CTCATTTTGGCTGTTGATCTGAGCAGTCTGACAGAAAGATATTACAAAGCTATTTTACC	1620

Query Match	100.0%;	Score	4177;	DB	4;	Length	4177;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	4177;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	ATGAGTAGTTTGGTGGAGAGCGCTTCGCATACGATCTGATAGGAACCCAGTTTATAACCTA	60				
DB	1	ATGAGTAGTTTGGTGGAGAGCGCTTCGCATACGATCTGATAGGAACCCAGTTTATAACCTA	60				
QY	61	GATGATTTCTGATGATGACGACTTGGTTCCTAAAAAAGATCGAACCTTTTGACGAAGTCGAG	120				
DB	61	GATGATTTCTGATGATGACGACTTGGTTCCTAAAAAAGATCGAACCTTTTGACGAAGTCGAG	120				
QY	121	GCTATTGTCAGAACTGATGGAAGAAATGCAATGTCAGGCTTGTGGGGAAAGTACTAAT	180				
DB	121	GCTATTGTCAGAACTGATGGAAGAAATGCAATGTCAGGCTTGTGGGGAAAGTACTAAT	180				
QY	181	CTTGTAAGCTGCAATACATGCACCTTATGCGTTCCATGCTAAATGCTTAGTTCCACCTCTT	240				
DB	181	CTTGTAAGCTGCAATACATGCACCTTATGCGTTCCATGCTAAATGCTTAGTTCCACCTCTT	240				
QY	241	AAAGATGCTTCCGTTGGAATAATGGAGATGCCCTGAATGTTGTTAGTCTCTTTAAACGAGATA	300				
DB	241	AAAGATGCTTCCGTTGGAATAATGGAGATGCCCTGAATGTTGTTAGTCTCTTTAAACGAGATA	300				
QY	301	GATGAATATTTGGATTTGTAATGCGTCTTCAAAAATCTAGTGAAACGAGTTCTCCGAT	360				
DB	301	GATGAATATTTGGATTTGTAATGCGTCTTCAAAAATCTAGTGAAACGAGTTCTCCGAT	360				
QY	361	GCGGAACCGAAGCCAAATTTTGTGAACAGTATCTCGTGAAGTGGAGGATTAATCATAC	420				
DB	361	GCGGAACCGAAGCCAAATTTTGTGAACAGTATCTCGTGAAGTGGAGGATTAATCATAC	420				
QY	421	CTTCACCTGCTCTTGGGTGCTCCTGAGAAGGATTCAGAAAGGCTTATAAGTCAAAATCATCGT	480				
DB	421	CTTCACCTGCTCTTGGGTGCTCCTGAGAAGGATTCAGAAAGGCTTATAAGTCAAAATCATCGT	480				
QY	481	TTAAAAACCGAGTGAAACAAATTTTACCGCTCAAAATGGAGTCAATTAACAGCGAAGAT	540				
DB	481	TTAAAAACCGAGTGAAACAAATTTTACCGCTCAAAATGGAGTCAATTAACAGCGAAGAT	540				

QY	1621	CGTAATATACAGTATTGACAAAAGGAGGGTCTCAAAATTTCCCTTAATAACATTATG	1680
Db	1621	CGTAATATACAGTATTGACAAAAGGAGGGTCTCAAAATTTCCCTTAATAACATTATG	1680
QY	1681	ATGGAATATACGAAAGTATGCTGCCATCTTATATGCTAGAGGGTCTGAGCCAGTTATT	1740
Db	1681	ATGGAATATACGAAAGTATGCTGCCATCTTATATGCTAGAGGGTCTGAGCCAGTTATT	1740
QY	1741	CACGACGCAAAATGAAGCTTTCAAAACAATTTTGGAGTCTTTGGAGCTGTTGGAAGCTGCAACTTTCTA	1800
Db	1741	CACGACGCAAAATGAAGCTTTCAAAACAATTTTGGAGTCTTTGGAGCTGTTGGAAGCTGCAACTTTCTA	1800
QY	1801	GATAAATATGATGCTCAAACTGAAAGAGCAAGGACACAGAGTCTTAATATACACAGTTT	1860
Db	1801	GATAAATATGATGCTCAAACTGAAAGAGCAAGGACACAGAGTCTTAATATACACAGTTT	1860
QY	1861	CAGCATATGCTGACCTTACTTGAAGACTACTGTACCATAAGAAATGGCAGTACGAGCGA	1920
Db	1861	CAGCATATGCTGACCTTACTTGAAGACTACTGTACCATAAGAAATGGCAGTACGAGCGA	1920
QY	1921	ATTGATGGAAGGTTGGCGAGCTGAGCGCAAAATACGATAGATCGGTTCAATGCCAAA	1980
Db	1921	ATTGATGGAAGGTTGGCGAGCTGAGCGCAAAATACGATAGATCGGTTCAATGCCAAA	1980
QY	1981	AATTTCTAAACAGTTTGTGCTCTCCACAGAGCTGGTGGCTTAGGAATAAATCTT	2040
Db	1981	AATTTCTAAACAGTTTGTGCTCTCCACAGAGCTGGTGGCTTAGGAATAAATCTT	2040
QY	2041	GCAACGGCTGATACAGTAATCATTTATGACAGTACTGGAATCCTCATGCTGATCTTCAA	2100
Db	2041	GCAACGGCTGATACAGTAATCATTTATGACAGTACTGGAATCCTCATGCTGATCTTCAA	2100
QY	2101	GCAATGGCTAGAGCTCATCGACTTGGCCAAAACAATAAGGTGATGATTTATAGGCTCAT	2160
Db	2101	GCAATGGCTAGAGCTCATCGACTTGGCCAAAACAATAAGGTGATGATTTATAGGCTCAT	2160
QY	2161	AACCGAGGACCAATGGAAGAGGATGATGCAATTTGACTGAAAGGAAATGGTTCTAGAG	2220
Db	2161	AACCGAGGACCAATGGAAGAGGATGATGCAATTTGACTGAAAGGAAATGGTTCTAGAG	2220
QY	2221	CATCTTTGTTGGGAACTCAAAACACAAAACATTAATCAGGAAGTTAGATGACATC	2280
Db	2221	CATCTTTGTTGGGAACTCAAAACACAAAACATTAATCAGGAAGTTAGATGACATC	2280
QY	2281	ATCAGGTATGGATCAAAAGGAGCTTTTGTCTAGTGAAGATGATGAACGAGAAAGTCTGGA	2340
Db	2281	ATCAGGTATGGATCAAAAGGAGCTTTTGTCTAGTGAAGATGATGAACGAGAAAGTCTGGA	2340
QY	2341	AAAAATTCATATGATGATCGGCTATAGACAAATTCGTTGATCGTATCTCGTGGAGGCA	2400
Db	2341	AAAAATTCATATGATGATCGGCTATAGACAAATTCGTTGATCGTATCTCGTGGAGGCA	2400
QY	2401	GAGGAAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTTCTAAAGGCTTTCAAGTGGCT	2460
Db	2401	GAGGAAGTCTCAGTGGATGATGAAGAGGAGAAATGGATTTCTAAAGGCTTTCAAGTGGCT	2460
QY	2461	AATTTTGAATATATAGATGAATGAGGACGACGATTAAGAGCAACAGAGAGTCTGCTGCT	2520
Db	2461	AATTTTGAATATATAGATGAATGAGGACGACGATTAAGAGCAACAGAGAGTCTGCTGCT	2520
QY	2521	GAAGCAAAATCTTCAGAGGCAATCTGATAGAGCAAGTTATTTGGGAAGAGTTGTTAAAA	2580
Db	2521	GAAGCAAAATCTTCAGAGGCAATCTGATAGAGCAAGTTATTTGGGAAGAGTTGTTAAAA	2580
QY	2581	GATAAATTTGAGCTGCACAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAGAGAAT	2640
Db	2581	GATAAATTTGAGCTGCACAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAGAGAAT	2640
QY	2641	CGCAAGCAGTTGTTATCCATTGAAGAGATGATCTTCTGCTGGTTTGGGAAGATGTGAGCTCT	2700
Db	2641	CGCAAGCAGTTGTTATCCATTGAAGAGATGATCTTCTGCTGGTTTGGGAAGATGTGAGCTCT	2700
QY	2701	GATGGAGTGAAGATTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAAGGAGTT	2760

Db	2701	GATGGAGTGAAGATTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGACAAGGAGTT	2760
QY	2761	CAGACGGGTCGACGGCGGTACAGAAAGGAGGTGCGGATAAATTTTGGAAACCAATCCGTTTG	2820
Db	2761	CAGACGGGTCGACGGCGGTACAGAAAGGAGGTGCGGATAAATTTTGGAAACCAATCCGTTTG	2820
QY	2821	ATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACACAGAGTCAAGGGGCAATT	2880
Db	2821	ATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACACAGAGTCAAGGGGCAATT	2880
QY	2881	TTTGTACAGACTTTGATGAGGTATGAGCTGCGCAATTTTGTGATTGGAAGAGTTTGTTCCT	2940
Db	2881	TTTGTACAGACTTTGATGAGGTATGAGCTGCGCAATTTTGTGATTGGAAGAGTTTGTTCCT	2940
QY	2941	CGCTTAAAGCAGAAGACCTTTTGAAGAAATAAATGAATATGGAATACTCTTTCTTGAAGCAC	3000
Db	2941	CGCTTAAAGCAGAAGACCTTTTGAAGAAATAAATGAATATGGAATACTCTTTCTTGAAGCAC	3000
QY	3001	ATTGCTGAAGAAATAGACGAGAAATCTCCAACTTTTCAGATGGTGTGCCCAAGGAAGGA	3060
Db	3001	ATTGCTGAAGAAATAGACGAGAAATCTCCAACTTTTCAGATGGTGTGCCCAAGGAAGGA	3060
QY	3061	CTTAGAATAGAAGATGTTCTAGTCAAGATTGCTCTTCTGATAGTCTTCTGAGGAAGGTTG	3120
Db	3061	CTTAGAATAGAAGATGTTCTAGTCAAGATTGCTCTTCTGATAGTCTTCTGAGGAAGGTTG	3120
QY	3121	AAATTTGTGAAGATCATCCAGGAAACCTGTTTCCCTCTCGCAATTTCTTGAAGATTC	3180
Db	3121	AAATTTGTGAAGATCATCCAGGAAACCTGTTTCCCTCTCGCAATTTCTTGAAGATTC	3180
QY	3181	CCCGACTGAGAAAGTGGAAAAATTTTGAAGAGGAAACATGACAAATATATATACGTGCT	3240
Db	3181	CCCGACTGAGAAAGTGGAAAAATTTTGAAGAGGAAACATGACAAATATATATACGTGCT	3240
QY	3241	GTTTTAAAGCATGGGTACGAGCGGTGGCAAGCTATTGTTGATGACAAAGATTGGGGATC	3300
Db	3241	GTTTTAAAGCATGGGTACGAGCGGTGGCAAGCTATTGTTGATGACAAAGATTGGGGATC	3300
QY	3301	CAAGAGCTTATCTGCAAAAGAAATTTGAATTTCCCTCACATAAGTTTGTCTGCTGAACAA	3360
Db	3301	CAAGAGCTTATCTGCAAAAGAAATTTGAATTTCCCTCACATAAGTTTGTCTGCTGAACAA	3360
QY	3361	GCTGTTTTCAGGGGAGAGTGGTATGGGGGCTTAATCCGGGAGCAGACATAACAG	3420
Db	3361	GCTGTTTTCAGGGGAGAGTGGTATGGGGGCTTAATCCGGGAGCAGACATAACAG	3420
QY	3421	AATCTGGAAGGTTTACTCGGACAAATAATGCTTCTGCTGATGGGCTCAAGTAAAC	3480
Db	3421	AATCTGGAAGGTTTACTCGGACAAATAATGCTTCTGCTGATGGGCTCAAGTAAAC	3480
QY	3481	TCGATGTTCTTATTCGGGACATGACAGACGACTTTGTTGAGTTTGTGAAAAAGCGAGTT	3540
Db	3481	TCGATGTTCTTATTCGGGACATGACAGACGACTTTGTTGAGTTTGTGAAAAAGCGAGTT	3540
QY	3541	CTGCTTTTGGAGAGGCGATGAATTAATGAAATGCTTGGTGAAGAAATATATGGAATTGTTGGC	3600
Db	3541	CTGCTTTTGGAGAGGCGATGAATTAATGAAATGCTTGGTGAAGAAATATATGGAATTGTTGGC	3600
QY	3601	TCATCATCTATCCCTACTGAAGAACAGAGCTGACCAAGATCGCTGACAGTGGGA	3660
Db	3601	TCATCATCTATCCCTACTGAAGAACAGAGCTGACCAAGATCGCTGACAGTGGGA	3660
QY	3661	GTGAGCTTATTTGAGGTTGATGATAATGCTTTGATGGAATCTTCTTAAGACTGATCTTATC	3720
Db	3661	GTGAGCTTATTTGAGGTTGATGATAATGCTTTGATGGAATCTTCTTAAGACTGATCTTATC	3720
QY	3721	ACTTCAGAGAAATATATGGGGCTGCTGTTGACAAACAAAGCGCGGCTCGAAATAGCT	3780
Db	3721	ACTTCAGAGAAATATATGGGGCTGCTGTTGACAAACAAAGCGCGGCTCGAAATAGCT	3780
QY	3781	CAACATTATACAGATGTGCAAACTTCTTGTGATGAGAACGCTCGGGAATCAGTCCAGCA	3840

Db 3781 CAACATTATAACCCAGATGTGCAAACTTCTTGATGAGAAAGCTCGGGAATCAGTCCAAGCA 3840
QY 3841 TATGTAACCAACCAACCCAGTACCAAGGTGAATGAGAGCTTCGGTGCACCTCAATCT 3900
Db 3841 TATGTAACCAACCAACCCAGTACCAAGGTGAATGAGAGCTTCGGTGCACCTCAATCT 3900
QY 3901 ATCAATGGTAACCAATTAACCAATCCCTTTTCGATTACATCTGATCAATCCCAAGTCACATGAA 3960
Db 3901 ATCAATGGTAACCAATTAACCAATCCCTTTTCGATTACATCTGATCAATCCCAAGTCACATGAA 3960
QY 3961 GACGACCAACCAAGCAGACCTTAACCAATGTTGAGATGAAGACACGCGCCGAAGAAACAAA 4020
Db 3961 GACGACCAACCAAGCAGACCTTAACCAATGTTGAGATGAAGACACGCGCCGAAGAAACAAA 4020
QY 4021 CCCTTAAGAGTGGCGTTCGATCTGAATGTTGGTGGAGGAGGACATTCGCTGA 4080
Db 4021 CCCTTAAGAGTGGCGTTCGATCTGAATGTTGGTGGAGGAGGACATTCGCTGA 4080
QY 4081 GCTAGTGGAGTGTGATGTAAATGGAAGAGCCAAAGAAAGAGAGCCAAAGAAC 4140
Db 4081 GCTAGTGGAGTGTGATGTAAATGGAAGAGCCAAAGAAAGAGAGCCAAAGAAC 4140
QY 4141 ATGTCGTTGATGACTCACTCACTCGTAAATCAAGATTC 4177
Db 4141 ATGTCGTTGATGACTCACTCGTAAATCAAGATTC 4177

RESULT 2

ABZ14721
ID ABZ14721 standard; DNA; 3204 BP.
AC ABZ14721;
XX
DT 21-JAN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 2526.
DE
XX Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX
XX WO200216655-A2.
XX
XX 28-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-US026685.
XX
XX 24-AUG-2000; 2000US-0227866P.
PR 26-JAN-2001; 2001US-0264647P.
PR 22-JUN-2001; 2001US-0300111P.
XX
PA (SCRI) SCRIPPS RES INST.
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
XX Harper JF, Kreps J, Wang X, Zhu T;
XX
XX WPI; 2002-304127/34.

Identifying a stress condition to which a plant cell has been exposed and producing plants with increased tolerance to these abiotic stresses.

Claim 144; SEQ ID NO 2526; 577pp + Sequence Listing; English.

The invention relates to identifying a stress condition to which a plant cell has been exposed, comprising: (a) contacting nucleic acid representative of expressed polynucleotides in the plant cell with an array or probes representative of the plant cell genome; and (b) detecting a profile of expressed polynucleotides in the plant cell characteristic of a stress response. The method is useful in the production of transgenic plants, cells and seeds and in producing plants with increased tolerance to abiotic stress. The present sequence is that of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used in methods of the invention. Note: The sequence data for this patent is

CC not represented in the printed specification but is based on sequence
CC information supplied to Derwent by the European Patent Office
XX
SQ Sequence 3204 BP; 1048 A; 561 C; 784 G; 811 T; 0 U; 0 Other;
Query Match 20.4%; Score 850.8; DB 6; Length 3204;
Best Local Similarity 61.0%; Pred. No. 8.2e-217;
Matches 1892; Conservative 0; Mismatches 697; Indels 513; Gaps 11;
QY 280 GTTAGTCTCTTAAACGAGATAGATATTCGATTTGGAATGCGTCTCAAAATCT 339
Db 121 GATTGTCCTCTTGGCGAGATTGAAAGATTTTGGACCGTGAATGGCGTCTACTGCATCT 180
QY 340 AGTGAACAAAGGTTCTCCCGAT---GGGAACCGAAGCCAATTTTGTGAAACAGATATCTC 396
Db 181 AACATCCGAATCTCTGACACGGGACACCAACCTAGTTGCTGTAACAAATATCTT 240
QY 397 GTGAAGTGAAGGATTTATCATCTTCACTGCTCTTGGGTGCTGAGAGGATTCAG 456
Db 241 GTGAAGTGAAGGATTTATCGTACTGCTCTTGGGTGCTGAGAGGATTCAG 300
QY 457 AGGCTTTAAGTCAATCATCTGTT-----TAAACACAGAGTGAAATTTTCAACCGT 510
Db 301 AAGGCTTACAGTCCCATCTCTTAAACATAAATAAGGGTGACCAAGTTTAAACGCT 360
QY 511 CAAATGGAGTCAATTC-----ATAACAGCAAGATGATTTTGTGCCATACGCTCTGAG 564
Db 361 GCTATGGAGTATTCATCGCTGAGAACGGAGCTCATGATTTATCGCCATTCGCTCGAA 420
QY 565 TGGACCACTGTTGATCGGATTTCTGCTGACAGAGGAAGATGGGAGCTGGAATATCTT 624
Db 421 TGGAAACCTGTTGACCGGATTTTGTCTGCCGGAAGGAGATGATGGGAGGAGTATCTG 480
QY 625 GTCAATATAAGAGCTATCTTATGATGATGTTTATGGAGTCAAGATCAGACATCTCA 684
Db 481 GTGAAGTACAAAGAACTTTCTTACAGAAATAGTACTGGGAATCTGAAATCTGACATCTCA 540
QY 685 ACCTTCCAGAAATGAAATTCAGAGGTTCAAGGATGTAATTTCTAGAACTCGCAGAAATAA 744
Db 541 GATTTCCAGAAATGAAATTCAGAGGTTTAAAGATATAAATCTAGCTCTCGTAGAGATAA 600
QY 745 GATGTTGACCAATAAAGAAATCCAGAGACTTTTCAACAGTTTATCATCTCTGAAATTC 804
Db 601 TATGTCGAAAAATGAGAGAAATCGAGAGGAATTCAAAACAGTTTGTATCTTACTCTGAAATTT 660
QY 805 CTCAAGGCTTGTATCATCATACAGCTTGAGGACTTAAATTTTTCGGGTTCTCGTG 864
Db 661 CTTACAGGCACATTCATACATACAGCTCGAAGGGCTGAAATTTCTTAAGGTATTCATGG 720
QY 865 TCAAAAACAGACGATGTAATCTCTGATGAAATGGGACTAGGCAAGACAAATTCAAAGC 924
Db 721 TCCAGAAACCAATGTAATCTCTGCGATGAAATGGGACTAGGCAAACTATTCAGAGC 780
QY 925 ATTGCCCCCTTTAGCTTTCACCTTTTGGAGGAACTCATTCGCGATTTGGTAAATGCTCT 984
Db 781 ATTGCCCCCTTTAGCTTTCACCTTTTGGAGGAACTCATTCCTCTCATCTAGTTGTTGCTCT 840
QY 985 CTATCACTCTGGTAACTGGGAGAGAGTTTGGCAGATTTGGCCCCCAGACAGATGAACGTG 1044
Db 841 CTCTCAACCATCGTAACTGGGAGAGGGAATTTGCCACCTGGGCAACACATATGAATGTG 900
QY 1045 GTTATGTTTGGCACTGGCAAGCTCGAGAGCTTATCAGAGAAATAGATGATTTTACTTA 1104
Db 901 GTTATGTAACCTGGGATTCGGAAGCACGATGATTTATGGAACATGATTTCTACTTC 960
QY 1105 TCGAAAGATCAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAG 1164
Db 961 TCTGAAGGGCGTAAAAAGGTAAAG----- 984
QY 1165 CAAAAAGATCAAGTTTGTATGTCCTCTCACATCGTATGATGATCAACCTAGATTCA 1224
Db 985 ----- 984

QY 1225 GCAGTTCTAAACCAATTAAAGTGGGAGTGCATGATTGTTGATGAAGGTCAATCGACTGAAA 1284
Db 985 ----- 984
QY 1285 AATAAGGATTCAAAGCTGTCTCTTCATTGACACAGATTTCAGTAACACCGTATTCCT 1344
Db 985 ----- 984
QY 1345 CTGACAGGAACACCACTTCAGAACAACTTGGATGAACCTTTTCATGCTCATGCAATTTCTTT 1404
Db 985 ----- 984
QY 1405 GATCGGGGAGTTTGGAAAGTTTGGAGAGTTCACAGGAGGTTCAAAGATATTATATCAA 1464
Db 985 ----- 984
QY 1465 GAGGAGCAGATCTCAAGGTTGCACAAAAATGTTGGCTCCACATTTGCTCAGAGGGTAAAA 1524
Db 985 -----AAA 987
QY 1525 AAAGCGTAAATGAAGACATGCCCCCAAAAGGAGCTCATTTTGGTGTGATCTGAGC 1584
Db 988 GACGTCTCAAGGATAAAGTGGCCCCCAAGAGGAGCTCATTTTGGCTGTGATATGAGC 1047
QY 1585 AGTCTCCAGAAAGAAATATTACAAGCTATTTTACCCTGTAATTTATCAAGTATTGACAAA 1644
Db 1048 AGCCAGCAAAAAGAGGTGTACAAGCTGTATTACCAATTAATTTATCAAGTCTTGACAAAG 1107
QY 1645 AAGGAGGTGCTCAAAATTTCCCTTAATAACATTATGATGAATTCAGAAAAGTATGCTGC 1704
Db 1108 AAAGGAGTGTAAATATTC-----TAAGTGTCTAATGAATTCAGCCAGTTTGTCTCT 1161
QY 1705 CATCTTTATATGCTAGAGGTTGTGAGCCAGTTATTACAGCCAAATGAAGCTTTCAAA 1764
Db 1162 CATCCGTATCTGTACAGACTTTGAGCCAAAGATTTGAAGAGCAAAATGAAGCTTTTACA 1221
QY 1765 CAACCTTTTGGAGTCTTGTGGAAGCTGCAACTTCTAGATAAATGATGTCAAACTGAAA 1824
Db 1222 AAACATTTGGAAGCTTTCCGAAAGTTGCAACTGTTAGACAAAATGATGTCAAACTCAAA 1281
QY 1825 GAGCAAGGACACAGAGTCTTAATATACACAGTCTTTCAGCATATGCTGGACTTTACTTGAA 1884
Db 1282 GAGCAGGGCACAGAGTGTAACTCTATACCAATTCACACACCCTTTACCTTTTGGAA 1341
QY 1885 GACTACTGTACCCATGAAGAAATGGCAGTACGAGCGAATTTGATGGAAGGTTGGCGAGCT 1944
Db 1342 GACTACTTCACTTTCAAGAAATTCGAATTTATGAACGGATTGATGGAAGATCAGTGGACCT 1401
QY 1945 GAGCGCAAAATACGCATAGATCGGTTCAATGCCAAAATTTCTAACAGTTTGTGTTTTTG 2004
Db 1402 GAGAGGCAAGTAAAGATAGATCGGTTTAAATGCAGAAAACTTCGAACCGGTTCTGTTTTTTG 1461
QY 2005 CTCTCCACAAGAGCTGTGGCTTTAGGAATAAATCTTTCACAGCTGATACAGTAATCATTT 2064
Db 1462 CTGAGTACAGAGCTGTGGTAATAGGATTAACCTTGGCAACGCTGATACAGTCATCATTT 1521
QY 2065 TATGACAGTCACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATCGACTT 2124
Db 1522 TATGACAGTGAATTTGGAATCCTCATGCTGATCTCCAGGCTATGCTAGAGTTTCATCGGCTT 1581
QY 2125 GGCACAAACAAATGAAGTATGATTATAGGCTCATAAACCGAGGACCATTTGAAGAAAGG 2184
Db 1582 GGCACAAACAAATGAAGTATGATCTACAGACTGATTCACAAAGGTACTGTAGAAGAACGG 1641
QY 2185 ATGATCAATTTGACTAAAAGAAAATGGTTCTAGAGCATCTTGTGTGGAAACTCAAA 2244
Db 1642 ATGATGAGATTAACCAAGAACAGATGCTTCTAGAGCATCTCGCTGTGGG-----A 1692
QY 2245 ACACAAAAATTAATCAGGAAGAGTTAGATGACATCATCAGGTATGGATCAAAAGGAGCTT 2304
Db 1693 AAACACATCTTTGTCCAGGACGAGTTAGATGACATCATCAATATATGTTCCAGGAACTT 1752
QY 2305 TTGCTAGTGAAGATGATGAAGCAGGAAAGTCTGGAAAAATTCATTATGATGATGCGGCT 2364

RESULT 3

Db 1753 TTCTCTGAAGAGAAATGATGAACAGGAGGTCTGGAAAAATCAATTTACATGATGCTGCT 1812
QY 2365 ATAGACAAATTCCTTGATCGTGTATCTCGTGGAGGAGAGAAAGTCTCAGTGGATGATGAA 2424
Db 1813 ATCGAACAGTCTGATAGTCAATCATGTAGATGCTGTGAGGCTCTCACTGGATGATGAG 1872
QY 2425 GAGGAGATGGAATCTTAAAGGCTTTCAAGGTGGCTTAATTTTGAATATATATAGTGAAT 2484
Db 1873 GAGGAACTGATTTTAAAGAAATTTCAAGGTGGCTAGTATTTTGAATGATGATGAA 1932
QY 2485 GAGGAGCAGCAGATTAGAGGCACAGAGTCTGCTGTAAGAGCAAAATCTTTCAGCAGCAAT 2544
Db 1933 AATGAGCTGACAGCATTTAGAGGAGCAAAAGTATAGAAAAACAATTTCTTCAGTAAGAAAT 1992
QY 2545 TCTGATAGACGAAGTTATTGGGAAGAGTTGTTTAAAGATATAATTTGAGCTGCAACGAGCT 2604
Db 1993 GCTGATAGAACAGTCACTCGAAAGATTTGCTGTAAGACAAATATGAAGTGCAACAAGCT 2052
QY 2605 GAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGAGTCCGCAAGCAGTTGTTATCCATTGAA 2664
Db 2053 GAGGAGCTCAGTGTCTCTTGGAAAGAGGAAAGAGAACCGCAAGCAGGTGATGTA---TGCT 2109
QY 2665 GAAGATGATCTTGTGTTTGGAGATGAGCTCTGATGGAGATGAAAGTTATGAAGCT 2724
Db 2110 GAAGATGATCTTGTGTTTGAAGAAATCTCTGATGAGGAGACGAATATTTGCTTGAT 2169
QY 2725 GAGTCAACA-----GATGGTGAAGCAGCAGGACAA 2754
Db 2170 GATTTAAAGTGACCTCTGATGAGGAGGAGGAGCAGATGAACCCGAGCAGCTCGACAG 2229
QY 2755 GGAGTTC---AGACGGGTGACGGCGTACAGAAAGAGGTCGCGATTAATTTTGGAAACCA 2811
Db 2230 AGAAAGCCGAGGACGCTTACAAGGCTTACAGAAAAAGGCTCGCGATTAATTTGCGAGGA 2289
QY 2812 ACTCCGTTGATGAGGAGGAGTCTTTCAGAGTACTGGTTTCAACAGAGTCAA 2871
Db 2290 ATACCTTTGATGGAAGCGAAGGAGATATCTCATGCTGCTGGATTTAACGAGACTGAG 2349
QY 2872 AGGGCCATTTTACAGACTTTGATGAGGTATGGAGCTGGCAATTTTGAATGGAAAGGAG 2931
Db 2350 AGGATATATTTTACGTACGTTTAAAGAGTATGGAATCTTATTTTGAACACATTTGCC 2409
QY 2932 TTTGTTCTCGCTTAAAGCAGAACCTTTTGAAGAAATAAATGAATATGGAATATCTCTTC 2991
Db 2410 GAAAAACCTTACCGATAATTTCAAAATTTTAAAGTTTATACGCAATGGT----- 2459
QY 2992 TTGAAGCACATTTGCTCAAGAAATAGACGAGAATTTCTCAACCTTTTCAGATGGTGGCCC 3051
Db 2460 -----CTATGCAATGGAGTTCCC 2478
QY 3052 AAGGAAGGACTTTAGAATAGAAGATGTTCTAGTCAGAAATTTGCTCTGATCTAGTTTCAG 3111
Db 2479 AAAGAAGGATTAAGCAGTACGAGCTACTAGTGAGCATGACTTTTATGATGCTAGTAAAG 2538
QY 3112 GAGAGGTGAATTTTGTAGAGATCATCCAGGGAACCTGTTTCCCTCTCCGATCTT 3171
Db 2539 GAGAAAGTGTCAATTTTGGACAAACCATCCGACAGCACCTGTTTT---CTTAACTACGTC 2595
QY 3172 GAAAGATTTCCCGACTGAGAAATGGAATAATTTGGAAGGAGAACATGCAAGATAATG 3231
Db 2596 ATCAGTAAGTACAATTTTGAGAAATGGAGCATTTTCTTAAGGAAGAACATGATAGGATCTG 2655
QY 3232 ATACGTGCTGTTTTTAAAGCATGGGTACGACCGGTGGCAAGCTATTGTTGATGACAAAGAG 3291
Db 2656 ATTCTGCTGTTTCCAAGCATGGCTATGATAGTGGTGGCCATCTGTTGAAGACGAAGAG 2715
QY 3292 TTGGGATCCAGAGCTTATCTGCAAGAAATTTGAATTTCCCT 3333
Db 2716 ATAGGTTCCAAGAGGTTGCTTGCCTGCAAGAGACCTGAATATCCCT 2757

```

AAT32301
ID AAT32301 standard; cDNA; 6327 BP.
AC AAT32301;
XX
XX 30-OCT-1996 (first entry)
DT
XX
DE Dermatomyositis specific autoantigen, Mi-2, coding sequence.
XX
XX Mi-2; autoantigen; collagen disease; chromosome 12; 12p13; helicase;
KW Dermatomyositis; diagnosis; ss.
XX
XX Homo sapiens.
XX
XX Location/Qualifiers
XX 1..5738
XX /tag= a
XX /product= "Mi-2"
XX /note= "the first ATG is at nucleotide 90, the sequence
XX given starts at nucleotide 91"
XX 1579..6417
XX /tag= b
XX polyA_signal 6234..6240
XX /tag= c
XX
XX DE19509279-C1.
XX
XX 15-MAY-1996.
XX
XX 15-MAR-1995; 95DE-01009279.
XX
XX 15-MAR-1995; 95DE-01009279.
XX
XX (PRIV-) PRIVATES INST IMMUNOLOGIE & MOLEKULARGEN.
XX
XX Seelig HP, Renz M;
XX
XX WPI; 1996-240280/25.
XX
XX P-PSDB; AAR99534.
XX
XX
XX DNA encoding dermatomyositis specific auto:antigen - useful for
XX differential diagnosis and treatment of dermatomyositis.
XX
XX Claim 3; Fig 2; 20pp; German.
XX
XX The present sequence encodes a 218 kD dermatomyositis specific auto-
XX antigen, designated Mi-2. The sequence numbering given in the
XX specification starts at nucleotide 91, i.e. the 5' UTR is omitted, as
XX well as the 'A' of the first ATG start codon. The protein is hydrophilic,
XX acidic and protruding regions of the protein are characteristic of
XX heliases. The gene corresponding to this cDNA was localised to
XX chromosome 12 (12p13). The DNA can be used for the recombinant production
XX of Mi-2 which is used for, e.g. the differential diagnosis of collagen
XX diseases, esp. dermatomyositis, e.g. by immunoassay or Western blotting
XX
XX Sequence 6327 BP; 1761 A; 1473 C; 1740 G; 1353 T; 0 U; 0 Other;
XX
XX Query Match 10.28; Score 424.6; DB 2; Length 6327;
XX Best Local Similarity 52.08; Pred. No. 1.9e-102;
XX Matches 1238; Conservative 0; Mismatches 1074; Indels 69; Gaps 10;
XX
XX 818 TACATCCATACACAGCTTGAGGAGCTTAATTTTGGCGTTCTCGTGGTCAAAACAGACGC 877
XX 2176 TGACCCCTCAATCAATGGAGGGCTGAATTTGGTTGGCTTCTCTGGGCTCAGGGCACTG 2235
XX
XX 878 ATGTAATCTTGTGATGAATGGAGTAGGCAAGACAATCAAGCAATTGCCCTTTTAG 937
XX 2236 ACACCATCTTGGCTGATGAGATGGGCTTGGGAAAACGTGTACAGACAGCAGTCTTCCTGT 2295
XX
XX 938 CTTCACTTTTGGAGG-----AACCTCATTCGGCATTTGGTAATTGCTCCTCTATCGA 991
XX 2296 ATTCCCTTTTCAAGAGGGGTCATTTCCAAAGGCCCTTCTTAGTAGGCGGCCCTCTTCTTA 2355

```

```

QY 992 CTCTCGGTAACCTGGGAGAGAGATTTGCCACATGGGCCCCACAGATGAACGTGGTTATGT 1051
DB 2356 CCATCATCAACTGGGAGCGGAGTTTGAATGTGGGCTCCAGACATGTATGCTGAACCT 2415
QY 1052 ATTTGGCACTGGCAAGCTCGAGGAGTTATCAGAGAACATGAGTTTACTTATCGAAAG 1111
DB 2416 ATGTGGGTGACAAGGACAGCGTGCATCATCCGAGAGAAATGAGTTTCTCTTTGAAGACA 2475
QY 1112 ATCAAAAAAGATCAAGAAAAAATCTGGACAAATAAGTAGCGGAAAAACAAGCAAAAAA 1171
DB 2476 ATGCCATTCGTGGTGCAGAGAGGCTC-----CCGATGAAGAAGAGGCAT 2523
QY 1172 GAATCAAGTTTGAATGTCCTCCTCAATCGTATGAGATGATCAACCTAGATTTCAAGAGTTC 1231
DB 2524 CTGTGAATTCATATGCTGCTGACATCCTATGAATGATCACCATTGACATGGCTATTT 2583
QY 1232 TAAACCAATTAAGTGGGAGTGCATGATTTGTGATGAAGGTGCATGACCTGAAAAATAAGG 1291
DB 2584 TGGGCTCTATTGATTTGGGCTGCCTCATCGTGGATGAAGCCCATCGGCTGAAGAACATC 2643
QY 1292 ATTCAAGCTGTTCTCTTCAATTGACACAGTATTCAAGTAACCAACCGTATTCTTCTGACAG 1351
DB 2644 AGTCTAAGTTCTTCGGGTATTGAATGGTTACTCACTCCAGCACAGCTGTTGCTGACTG 2703
QY 1352 GAACACCACTTCAGAACAACTTTGGATGAACCTTTTCAATGCTCATGCAATTTCTTGATCGGG 1411
DB 2704 GGACACCATTAACAAAAAATCTGGAAGAGTTGTTTCACTGCTCAACTTTCTCACCCCG 2763
QY 1412 GGAAGTTTGGAGTTTGGAGGAGTTCCAGGAGGAGTTCAAGATATTAATCAAGAGGAGC 1471
DB 2764 AGAGGTTCCACAATTTGGAAGGTTTGTGGAGAGTTTCTGACATGCGCAAGGAGGACC 2823
QY 1472 AGATCTCAAGGTTGCACAAAAATTTGGCTCCACATTTGCTCAGAAAGGGTAAAAAAGAGC 1531
DB 2824 AGATAAAAAAATCTGCATGACATGCTGGGCGCGCACATGTTGCGCGGCTCAAAGCCGAGC 2883
QY 1532 TAATGAAGACATGCCCCCAAAAAAGAGCTCATTTTGGGTGTGATCTGAGAGAGTCTGC 1591
DB 2884 TGTTCAGAAACATGCCCTCCAAGACAGAACTAATTTGTGGGTGTGGAGCTGAGCCCTATGC 2943
QY 1592 AGAAGAAATTAACAAAGCTATTTTACCGTAATTAATCAAGTATTTGA---CAAAAAAGG 1648
DB 2944 AGAAGAAATTAACAAAGCTATTTTACCGTAATTTTGAAGCACTCAATGCCCGAGGTG 3003
QY 1649 GAGGTGCTCAAAATTTCCCTTAATAACATTAATGATGAATTTACGAAAGATGCTGCGCATC 1708
DB 3004 GTGGCAACAGGTGCTCTGCTGAATGGTGTGATGATCTTAAGAAGTGTGCAACCATC 3063
QY 1709 CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTACGACGCAAAATGAAGCTT 1759
DB 3064 CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCCTAAGATGCCATAATGGCATGTATGATG 3123
QY 1760 TCNAACAACTTTGGAGTCTTGTGGAAGCTGCAACTTCTAGATAAAATGATGGTCAAAAC 1819
DB 3124 GCAGTCCCTTAATCAGAGCATCTGGGAAATATTGCTGCTGCAGAAAAATGCTCAAGAAC 3183
QY 1820 TGAAGAGCAAGGACACAGAGTCTTAATATACACACAGTTTTCAGCATATGCTGGCATTC 1879
DB 3184 TTAAGAGGGTGGCATGTGTACTCATCTTTCCAGATGACCAAGATGTAGACCTGC 3243
QY 1880 TTGAAGACTACTGTATCCCATTAAGAAATGGCAGTACGAGGCAATGATGGAAGAGTTGGCG 1939
DB 3244 TAGAGGATTTCTTGGAAACATGAAGTTATAAAATACGAAGCATCGATGGTGGAACTCACTG 3303
QY 1940 GAGCTGAGGGGCAAAATACGATAGATCGGTTCATGTCGCAAAAAATTTCTAACAATTTGTT 1999
DB 3304 GGAACATCGGGCAAGAGGCCATTTGACCGCTTCAATGCACCGGGTGTCTCAGAGTTCTGCT 3363
QY 2000 TTTTGTCTCCCAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA 2059
DB 3364 TCTTGTCTTCCATCTCAGCTGGGGGCTTGGAAATCAATCTGGCCACTGCTGACACAGTTA 3423
QY 2060 TCATTTATGACAGTGAAGTGAATTCCTCATGCTGATCTTCAAGCAATGGGCTAGAGCTCATC 2119

```


Db 3424 TTATCTATGACTCTGACTGGAACCCCAATAATGACATTCAGGCCCTTTAGCAGAGCTCACC 3483
Qy 2120 GACTTGGCCAAACAAATAAGTGTATGATTTATAGGCTCATAAACCGAGGACCACTTGAAG 2179
Db 3484 GGATTGGCAAAATAAAAGTAAATGATCTACCGTTTGTGACCCGTCGTGAGTGAGG 3543
Qy 2180 AAAGGATGATGCAATTGACTTAAAGAAATGTTCTAGAGCATCTTTGTTGG----- 2234
Db 3544 AGCGCATGACGAGTGGCAAGAAAGAAATGATGCTGACGCATCTAGTGGTGGCGCTG 3603
Qy 2235 -----GAAACTCAAAACACAAACATTAATCAGAGAGTTAGATGACATCATCAGGTATG 2290
Db 3604 GGTTGGGCTCAAGACTGGATCTATGTCCAACAGGAGCTTGAATGATATCCTCAAAATTG 3663
Qy 2291 GATCAAGAGGAGCTTTTGTCTAGTCAAGATGATGAAGCAGGAAAGTCTGGAAAAATTCATT 2350
Db 3664 GCACTGAGGAACATTAACAAGATGAGCCACTGATGGAGGAGGACAAACAGAGGGAG 3723
Qy 2351 ATGATGATGGGCTATATAGACAAATTTGCTGA-----TCGTGATCTGTGGAGGCGAGG 2404
Db 3724 AAGATGACAGTGTATCCACTAGCATGAAGGCCATTTGAACGGCTGTAGACCGTAACC 3783
Qy 2405 AAGTCTCAGTGGATGATGAAGAGGAGATGATCTTTAAAGGCTTTCAAGGTGGCTAATT 2464
Db 3784 AGATGAGACTGAAGACACAGAAATTCAGGGCATGAATGAATATTTGAGCTCAATCAAAG 3843
Qy 2465 TTCAATATATAGATGAAATGAGGACGACGACATTTAGAGGCACAGAGATCGCTGTGAAA 2524
Db 3844 TGCCCCATGATGTGTACGGGAGAGAAATGGGGAGGAGGACGTAGAACGGGAAA 3903
Qy 2525 GCAAAATCTTCAGCAGGCAATTTGATAGACAAAGTTATTCGGAGAGTGTGTTAAAAAGATA 2584
Db 3904 TCATTAACAGGAAGAAAGTGTGATCCTGA---CTACTGGGAGAAATTCCTGGGCACC 3960
Qy 2585 AATTGAGCTGCACCGCTGAGGAGCTTAATGCTCTTGGAAGAAAGAGAGAGTCGCA 2644
Db 3961 ATTATGAGCAGCAAGCAAGATCTAGCCGAAATCTGGGCAAGGAAAGAAATCCGTA 4020
Qy 2645 AGCAGTTGGTATCATTGAAG---AAGATGATCTGTGCTGTTGGAGATGTGAGCTCTG 2701
Db 4021 AACAGTCAACTACATGATGCTGCCAGAGGACCGAGATTTGCAGGACGACAGTCCG 4080
Qy 2702 ATGAGATGAAGATTATGAGCTGAGTCAACAGATGTTGAAGCAGCAGGACAGGAGTTC 2761
Db 4081 ACAACAGTCCGATTACTCAGTGGCTTCAGAGGAAGTGATGAAGACTTTGATGAACGTT 4140
Qy 2762 AGCGGTTCACGCCGTACAGAAAGAGGTGCGGATATTTGGAAACCAA----- 2812
Db 4141 CAGAAGCTCCCGTAGGCCAGTCTGTAAGGGCCCTGGCGGAATGATAAAGATAAGCCATTGC 4200
Qy 2813 CTCGTTGATGGAAGGTGAGGGAGATCTTTTCAGAGTACTGGGTTTCAACGAGAGTCAAA 2872
Db 4201 CTCTCTGTGGCCGTTGTTGGGAAATATGAAGTACTTGGTTTAAATGCTCGTCAGC 4260
Qy 2873 GGCCCAATTTTGTACAGACTTTGATGAGGTATG------AGCTGGCAATTTGATT 2923
Db 4261 GAAAAGCCTTTCTTAATGCAATATATGATGATGATGTCACCTCAGGATGCTTTTACTA 4320
Qy 2924 GGAAGAGTTTGTCTCGCTTAAAGCAGAGACCTTTTGAAGAAATAAATGAATGGA 2983
Db 4321 CCCAGTGGCTTGAAGAGACCTCGAGGCAATCAGAGAAAGAGTTCAAGGCATATGTCT 4380
Qy 2984 TACTCTTTGAGACCATTTGTAAGAAATAGACGAGAATTTCTCAACCTTTTTCAGATG 3043
Db 4381 CTCTTTTCATCGGCATTTATGTAGCCGGGGCAGATGGGGCTGAGACCTTTGCTGATG 4440
Qy 3044 GTGTGCCCAAGGAGGACTTAGAATAGAAGATGTTCTTAGTCAAGATTTGCTCTTGATAC 3103
Db 4441 GTGTCCCGGAGAGGCTGTCTCGCCAGCATGCTCTTACTAGAATTTGGTGTATGCTT 4500
Qy 3104 TAGTTAGGAGAGGTTGAAATTTGTAAGCATATCCAGG 3144

Db 4501 TGATTCGCAAGAGGTTTCAGGATTTGAACATGTTAATGGG 4541
RESULT 4
ABK70271
ID ABK70271 standard; cDNA; 5739 BP.
XX ABK70271;
XX 15-JUL-2002 (first entry)
XX Human lung cancer associated full length cDNA DMSM-6.
DE Human; ss; gene; lung cancer; cytostatic; tumour; vaccine.
XX Homo sapiens.
XX WO200224057-A2.
XX 28-MAR-2002.
XX 20-SEP-2001; 2001WO-US042232.
XX 22-SEP-2000; 2000US-0234837P.
XX 10-OCT-2000; 2000US-0239440P.
XX 29-JUN-2001; 2001US-0301928P.
XX (CORI-) CORIXA CORP.
XX Benson DR, Mohamath R, Lodes MJ;
XX WPI; 2002-372001/40.
XX New tumor lung proteins and nucleic acids encoding the proteins, useful
XX as vaccines and for treating, preventing, diagnosing or monitoring lung
XX cancer.
XX Claim 1; Page 146-148; 189pp; English.
XX The invention relates to an isolated polynucleotide comprising a sequence
XX selected from 183 human DNA sequences (appearing as ABK70130-ABK70312),
XX or their fragments, homologues, variants or complements and their encoded
XX polypeptides. Also included are an expression vector comprising the
XX polynucleotide operably linked to an expression control sequence; a host
XX cell transformed or transfected with an expression vector of; an isolated
XX antibody, or its antigen-binding fragment that specifically binds to the
XX polypeptide; a method for detecting the presence of a cancer in a patient
XX; a fusion protein comprising at least the polypeptide; an
XX oligonucleotide that hybridises to the polynucleotide under moderately
XX stringent conditions; a method for stimulating and/or expanding T cells
XX specific for a tumour protein; an isolated T cell population comprising T
XX cells prepared from the method of above; a composition comprising a first
XX component consisting of carriers and immunostimulants, and a second
XX component selected from the polynucleotides, proteins, antibodies, fusion
XX proteins, T cell populations and antigen presenting cells expressing the
XX polypeptide; methods for stimulating an immune response or treating
XX cancer in a patient by administering the composition and diagnostic kits
XX comprising at least one of the oligonucleotide of, or an antibody and a
XX detection reagent consisting of a reporter group. The polypeptides and
XX polynucleotides are useful as vaccines for the treatment or prevention of
XX lung cancer, and for diagnosis and monitoring of such cancer. The
XX polynucleotide, polypeptide and antigen presenting cells can be used to
XX stimulate or expand T cells specific for a tumorous protein. The
XX polynucleotides may be used as probes or primers for nucleic acid
XX hybridisation, and in the preparation of ribozyme molecules for
XX inhibiting expression of tumour polypeptides and proteins in tumour
XX cells. The present sequence is one of the 183 lung cancer associated
XX polynucleotides
SQ Sequence 5739 BP; 1618 A; 1323 C; 1601 G; 1197 T; 0 U; 0 Other;

Query Match 10.1%; Score 421.4; DB 6; Length 5739;
Best Local Similarity 51.9%; Pred. No. 1.3e-101;


```
Qy 2924 GGAGGAGTTTGTCTCCTCGCTTAAGCAGAGACCTTTGAAGAAATAAATGAATATGAGAA 2983
Db 4322 CCAGTGGCTTGAAGAGACCTCGAGGCAAAATCAGAGAAAGAGTTCAAGGCATATGTCT 4381
Qy 2984 TACTCTTTTGAAGACCAATTCCTGAAGAAATAGACGAGAAATCTCCAAACCTTTTTCAGATG 3043
Db 4382 CTCTTTTCATCGGCATTTATGTGAGCCGGGGCAGATGGGCTGAGACCTTTTCTGTGATG 4441
Qy 3044 GTGTGCCAAGGAGGACTTAGAATAGAGATGTTCTAGTCAGAAATGCTCTTCTGATAC 3103
Db 4442 GTGTCCCGCAGAGAGGCTGCTCGCCAGCATGCTCTTACTAGAAATGGTGTATGTCTT 4501
Qy 3104 TACTTCAGGAGAGGTTGAAATTTGTAGAAGATCATCCAGG 3144
Db 4502 TGATTCGAAGAAGGTTTCAGAGTTTGAACATGTTAATGGG 4542

RESULT 5
ABL64410
ID ABL64410 standard; DNA; 6417 BP.
XX AC ABL64410;
XX AC
XX 15-MAY-2002 (first entry)
XX DE Stomach cancer related gene sequence SEQ ID NO:2747.
XX DE
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX OS Homo sapiens.
XX PN WC200194629-A2.
XX PD 13-DEC-2001.
XX PF 30-MAY-2001; 2001WO-US010838.
XX PR 05-JUN-2000; 2000US-0209473P.
XX PR 08-JUN-2000; 2000US-0209531P.
XX PR 18-SEP-2000; 2000US-0233133P.
XX PR 18-SEP-2000; 2000US-0233133P.
XX PR 20-SEP-2000; 2000US-0234009P.
XX PR 20-SEP-2000; 2000US-0234034P.
XX PR 20-SEP-2000; 2000US-0234052P.
XX PR 22-SEP-2000; 2000US-0234509P.
XX PR 22-SEP-2000; 2000US-0234567P.
XX PR 25-SEP-2000; 2000US-0234923P.
XX PR 25-SEP-2000; 2000US-0234924P.
XX PR 25-SEP-2000; 2000US-0235077P.
XX PR 25-SEP-2000; 2000US-0235082P.
XX PR 25-SEP-2000; 2000US-0235134P.
XX PR 26-SEP-2000; 2000US-0235280P.
XX PR 26-SEP-2000; 2000US-0235637P.
XX PR 27-SEP-2000; 2000US-0235711P.
XX PR 27-SEP-2000; 2000US-0235720P.
XX PR 27-SEP-2000; 2000US-0235740P.
XX PR 27-SEP-2000; 2000US-0235840P.
XX PR 28-SEP-2000; 2000US-0236028P.
XX PR 28-SEP-2000; 2000US-0236032P.
XX PR 28-SEP-2000; 2000US-0236033P.
XX PR 28-SEP-2000; 2000US-0236034P.
XX PR 28-SEP-2000; 2000US-0236109P.
XX PR 28-SEP-2000; 2000US-0236111P.
XX PR 29-SEP-2000; 2000US-0236842P.
XX PR 29-SEP-2000; 2000US-0236891P.
XX PR 02-OCT-2000; 2000US-0237172P.
XX PR 02-OCT-2000; 2000US-0237173P.
XX PR 02-OCT-2000; 2000US-0237278P.
XX PR 02-OCT-2000; 2000US-0237294P.
```

```
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX (AVAL-) AVALON PHARM.
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Sopet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX Claim 1; SEQ ID NO 2747; 44pp; English.
XX The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour
XX SQ Sequence 6417 BP; 1774 A; 1494 C; 1784 G; 1365 T; 0 U; 0 Other;
Query Match 10.1%; Score 421.4; DB 6; Length 6417;
Best Local Similarity 51.9%; Pred. No. 1.4e-101;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;
Qy 818 TACATCCATACCAGCTTGAGGGACTTAATTTTGGGTTTCTCGTGTCAAACACAGCGC 877
Db 2266 TGCACCCCTATCAATGGAGGGCTTGAATGGTTGGCTTCTCTGGGCTCAGGCACTG 2325
Qy 878 ATGTAATCCTTGTGATGAAATGGGACTAGGCAAGACAAATTCAAAGCATTGCCCTTTTAG 937
Db 2326 ACACCATCTTGGCTGATGAGATGGGCTTGGGAAACTGTACAGACAGCAGTCTTCTGT 2385
Qy 938 CTTCACTTTTGAAGGAG-----AACCTCATTCGGCATTGGTAATGTCCTCTATCGA 991
Db 2386 ATTCCCTTTACAGGAGGGTTCATTCCAAAGGCCCTTCTCTAGTAGGCGCCCTTTCTA 2445
Qy 992 CTCTGCGTAACGGGAGAGAGATTTGCCACATGGGCCCCACAGATGAAGCTGTTATGT 1051
Db 2446 CCATCATCACTGGGAGCGGAGTTTGAATGTGGGCTCGAGCATGTATGTCTAACCT 2505
Qy 1052 ATTTTGGCCTCGCAAGCTCGAGCAGTTTATCAGAGAACATGAGTTTTTACTTATCGAAAG 1111
Db 2506 ATGTGGTGCACAGGACAGCCGTGCCCATCATCCGAGAGATGAGTTCTCTTTGAAGACA 2565
Qy 1112 ATCAAAAAAGATCAAGAAAAAGAAATCTGGCAAAATAAGTAGCGAAAGACACAAAAA 1171
Db 2566 ATGCCATTGCTGTGGCAAGAGGCCCTC-----CCGCATGAAGAAAGAGGCAT 2613
Qy 1172 GAATCAAGTTTGTGTCCTCCCTCACATCGTATGAGATGATCAACCTAGATTTCAGCAGTTC 1231
```

Db 2614 CTGTGAATTCATCTGCTGACATCCTATGAATTGATCACCAATTGACATGCTATTT 2673
 Qy 1232 TAAACCAATTAAGTGGAGTGCATGATGTTGATGAAGTGCATGACTGAATTAAGG 1291
 Db 2674 TGGGCTCTATTGATTGGGCTCTCATCGTGGATGAAGCCATCGGCTGAAGAACATC 2733
 Qy 1292 ATTCAAAGCTGTTCTCTTCAATGACACAGTATTCAAGTAACACCGTATTCTTCTGACAG 1351
 Db 2734 AGTCTAAGTCTTCCGGGTATTGAATGGTACTCACTCCAGCACAAAGCTGTTGCTGACTG 2793
 Qy 1352 GAACCACTTCAAGAACTTGGATGAACATTTTCACTGCTCATGCTTCTTCTGATCGG 1411
 Db 2794 GGACACATTACAAAACAATCTGAAGAGTGTGTTCTATCTGCTCAACTTTCTCACCCCG 2853
 Qy 1412 GGAAGTTTGGAGTGTGGAGAGTTCAGAGGAGTTCAAAGATATTAAATCAAGAGAGC 1471
 Db 2854 AGAGGTTCCCAATTTTGAAGGTTTTTGGAGGAGTTGCTGACATTTGCCAAGAGAGACC 2913
 Qy 1472 AGATCTCAAGGTTGCACAAAATGTTGGCTCCACATTTGCTCAGAAAGGTAAAAAAGAGC 1531
 Db 2914 AGATAAAAAAACTGCATGACATGCTGGGCGGCACATGTTGCGGCGCTCNAAGCCGATG 2973
 Qy 1532 TAATGAAGAAGATGCCCCCAAAAGAGCTCATTTTGGCGTGTGATCTGAGCAGTCTGC 1591
 Db 2974 TGTTCAGAAACATGCCCCCTCCAAGACAGAACTAAATTGTGCGTGTGGAGCTGAGCCCTATGC 3033
 Qy 1592 AGAAGATATTACAAAGCTATTTTACCCTTAATTATCAAGTATTGA---CAAAAAGG 1648
 Db 3034 AGAAGAAATCTCAAGATPACATCTCACTCGAATTTTGAAGCACTCAATGCCGAGGTG 3093
 Qy 1649 GAGGTGCTCAAAATTTCCCTTAATAACATTAATGATGAATTAACGAAAGTATGCTGCCATC 1708
 Db 3094 GTGGCAACCAAGTGTCTCTGCTGATGATGTTGATGATCTTAAAGATGCTGCAACCATC 3153
 Qy 1709 CTTA-----TATGCTAGAGGTTGTGAGCAGTATTATCACAGCAAAATGAAGCTT 1759
 Db 3154 CATACCTCTTCCCTGTGCTGCAATGGAAGCTCTCAAGATCCCTAATGGCATGTATGATG 3213
 Qy 1760 TCAACAACATTTTGGAGTCTTGGAGAGTGCACACTTCTAGATAAATGATGCTCAAC 1819
 Db 3214 GCAGTGCCCTAATCAGAGCATCTGGGAAATATTGCTGCTGCAGAAATGCTCAAGAAC 3273
 Qy 1820 TGAAGAGCAAGGACACAGAGTCTCTAATATACACACAGTTCACAGCATATGCTGCACTTAC 1879
 Db 3274 TTAAGGAGGTGGGATCGTACTCACTTTCCAGATGACCAAGATGCTAGACCTGC 3333
 Qy 1880 TTGAAGACTACTGTACCCATGAAGAAATGGCAGTACGAGCGAATTTGATGGAAGGTTGGCG 1939
 Db 3334 TAGAGGATTTCTTGGAAATGAAGGTTATAAATACGAACGCATCGATGGTGGAAATCACTG 3393
 Qy 1940 GAGCTGAGCGCAATACGCATAGATCGGTTCAATGCCAAAATTTCTAACAAGTTTGTGTT 1999
 Db 3394 GGAAATCGGCAAGAGGCCATTTGACCGCTTCAATGCAACCGGGTGTCTCAGCAGTTCGTCT 3453
 Qy 2000 TTTTGTCTCTCCCAAGAGCTGGTGGCTTAGGAATAAATCTTGGCAACGGCTCATACAGTAA 2059
 Db 3454 TCTTGTCTTCCACTCGAGCTGGGGCTTGGATCAATCTGCGCACTGCTGACACAGTTA 3513
 Qy 2060 TCATTTATGACGTGACTGGAATCTCTCATGCTGATCTTTCAAGCAATGGCTAGAGCTCATC 2119
 Db 3514 TTATCTATGACTCTGACTGGAACCCCAATAATGACATTTCAAGGCTTTAGCAGAGCTCAC 3573
 Qy 2120 GACTTGGCCAAACAAATAGGTGATGATTTATAGGCTCATAAACCGAGGCCACCATGAG 2179
 Db 3574 GGAITGGGCAAAATAAAGGTAATGATCTACCGGTTTGTGACCGCTGCGTCACTGGAGG 3633
 Qy 2180 AAAGGATGATGCAATTTGACTAAAAAGAAAAATGGTTCTTAGACATCTTGTGTTGG----- 2234
 Db 3634 AGCGCATCACGAGGTGGCAAGAGAAATGATGCTGACGCACTAGTGGTGGGCGCTG 3693
 Qy 2235 ----GAAACTCAAAACACAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290
 Db 3694 GGCTGGGCTCCAAGACTGGATCTATGTCCAAACAGGAGCTTTGATGATATCTCTCAAAATTTG 3753

Qy 2291 GATCAAAGGAGCTTTTGTGCTAGTGAAGATGATGAAGCAGGAAAGTCTGGAAAAATTCATT 2350
 Db 3754 GCACTGAGGAACTATTCAAGATGAAGCCACTGATGGAGGAGGAGACAACAAGAGGGAG 3813
 Qy 2351 ATGATGATGCGGTATAGACAAATTCGTTGA-----TCGTGATCTGCTGGAGGAGAGG 2404
 Db 3814 AAGATAGCAGTGTATCCACTACGATGATAAGCCATTGAACCGCTGCTAGACCGTAACC 3873
 Qy 2405 AAGTCTCAGTGGATGATGAAGAGGAGATGATCTTTAAAGGCTTTCAAGGTGGCTAATT 2464
 Db 3874 AGATGAGACTGAAGACACAGAAATTCAGGGCATTGAATGATATTGAGCTCATTTCAAAG 3933
 Qy 2465 TTCAATATATAGATGAAAATGAGGACAGCAGCAATTAGAGGCACAGAGAGTCCGTCTGAAA 2524
 Db 3934 TGCCCAAGTATGTTGATCGGAGAGAAATGCGGAGGAGAGAGGAGGTAGAACGGGAAA 3993
 Qy 2525 GCAAAATCTTCAGCAGGCAATTTCTGATAGACAAAGTATTATGGGAAGAGTTCGTTAAAAGATA 2584
 Db 3994 TCATTTAAACAGGAAGAAAGTGTGATCTCTGA---CTACTGGAGAAAATTCGTCGGCACC 4050
 Qy 2585 AATTGAGCTGCACCGGCTGAGGAGCTTAATGCTCTTGGAAAAGGAGAGAGTCCGCA 2644
 Db 4051 ATTATGAGCAGCAGCAAGAAAGATCTAGCCGAAATCTGGGCAAGGAAAAGAAATCCGTA 4110
 Qy 2645 AGCAGTTGGTATCCATTGAAAG---AAGATGATCTTGTGTTTGGAAAGATGTGAGCTCTG 2701
 Db 4111 AACAGGTCACACTACATGATGGCTCCAGGAGGACCGAGATTGCGAGGACGACAGTCCG 4170
 Qy 2702 ATGGAGATGAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGAGAGGAGTTC 2761
 Db 4171 ACAACAGTCCGATTAATCTAGTGGCTTTCAGAGGAAGTATGAAGACTTTTGATGAACGTT 4230
 Qy 2762 AGAGGCTCAGCGGCTCAGAAAGAGGTCGCGATATATTTGGAAACCA----- 2812
 Db 4231 CAGAAGCTCCCCGTAGGCCCAGTCGTAAGGGGCTTCGGAATGATAAAGATAAGCCATTGC 4290
 Qy 2813 CTCGTTGATGGAAGGTGAGGAGAGATCTTTTCAGAGTACTTGGGTTTCAACCCAGAGTCAAA 2872
 Db 4291 CTCTCTGTTGGCCGCTGTTGGGGAATATTGAAGTACTTGGTTTAAATGCTCGTCAGC 4350
 Qy 2873 GGCCCAATTTTGTACAGACTTTGATGAGGTATGG-----AGCTGGCAATTTTGTATT 2923
 Db 4351 GAAAAGCTTTCTTAATGCAATTTATGCAATATGATGATGCCACCTCAGGATGCTTTACTA 4410
 Qy 2924 GGAAGGATTTGTTCTCGCTTAAAGCAGAGAGACCTTTTGAAGAAATAAATGAATATGAA 2983
 Db 4411 CCAGTGGCTTGTAAAGAGACCTTCGAGGCAATCAGAGAAAGAGTTCAGGCAATATGCT 4470
 Qy 2984 TACTCTTTTGAAGCAATTTGCTGAAGAAATAGACGAGAAATTTCTCAACCTTTTCAGATG 3043
 Db 4471 CTCTTTTCATGCGGCAATTTATGAGCCGGGCGAGATGGGCTGAGACCTTTTGTGATG 4530
 Qy 3044 GTGTGCCAAAGGAAGCACTTAGAATAGAGATGTTCTAGTCAAGATTTGCTTCTGTATAC 3103
 Db 4531 GTGTCCCGGAGAGGCTGTCTCGCCAGCATGCTCTTACTAGAAATGGTGTATGTCTT 4590
 Qy 3104 TAGTTCAGGAGAGGTGAAATTTGTAGAAGATCATCCAGG 3144
 Db 4591 TGATTCCGAAGAAGGTTTCAGGAGTTTGAACATGTTAATGGG 4631

RESULT 6

ABX04169
 ID ABX04169 standard; cdna; 6417 BP.

XX ABX04169;

XX AC
 XX AC
 DT 10-JAN-2003 (first entry)

XX Human mRNA differentially expressed in mesenchymal cells #16.

DE Human; ss; gene; skeletal growth; cartilage degeneration disorder;

XX Human; ss; gene; skeletal growth; cartilage degeneration disorder;

KW Human; ss; gene; skeletal growth; cartilage degeneration disorder;

KW	chondroblastic phenotype; mesenchymal cell; cartilage formation; bone formation; arthritis; osteoarthritis; rheumatoid arthritis; gout arthritis; adjuvant arthritis; arthritis deformans; anti-gout; infectious arthritis; osteochondrosis; RDA; antiarthritis; osteopathic; antirheumatic; antiinflammatory; representational difference analysis.
OS	Homo sapiens.
XX	WO200271927-A2.
PN	19-SEP-2002.
PD	12-MAR-2002; 2002WO-US007787.
XX	12-MAR-2001; 2001US-0274980P.
PF	(BGHM) BRIGHAM & WOMENS HOSPITAL INC.
XX	Yates KE, Mizuno S, Glowacki J;
PI	WPI; 2002-732276/78.
XX	New nucleic acid molecules capable of promoting chondrogenesis, useful for diagnosing and treating cartilaginous tissue degeneration conditions, e.g. osteoarthritis, rheumatoid arthritis, gout arthritis, or osteochondrosis.
PT	Claim 33; Page 111-113; 153pp; English.
XX	The invention relates to new isolated nucleic acid molecule comprising a nucleic acid molecule consisting of a gene differentially expressed in cells undergoing differentiation from mesenchymal cell to a chondroblastic phenotype, or hybridising under stringent conditions to them (or their fragments). Also included are expression vectors, transformed host cells, expressed polypeptides or peptide fragments (which induce differentiation of a mesenchymal cell and may be used as an immunogen), binding partners of the polypeptides, a method for identifying an agent useful in modulating mesenchymal cell differentiation induction activity of a molecule, a method of diagnosing a condition characterized by aberrant expression of a nucleic acid molecule or its expression product; a method for determining regression, progression or onset of cartilaginous tissue degeneration condition in a subject characterised by aberrant expression of a nucleic acid molecule or its expression product, a method for treating a cartilaginous tissue degeneration condition, a method for treating a subject to reduce the risk of cartilaginous tissue degeneration condition developing in the subject, a method for identifying a candidate agent for treating a cartilaginous tissue degeneration condition, and a solid-phase nucleic acid molecule array consisting essentially of a set of nucleic acid molecule as cited above (or known from known genes shown to be differentially expressed in developing mesenchymal cells using the technique of representational difference analysis, RDA), its expression products or fragments, fixed to a solid substrate. The nucleic acids, polypeptides and agents are useful for treating cartilaginous tissue degeneration conditions such as osteoarthritis, rheumatoid arthritis, gout arthritis, adjuvant arthritis, arthritis deformans, infectious arthritis or osteochondrosis. The present sequence is a cDNA from a known gene differentially expressed in developing mesenchymal cells
XX	Sequence 6417 BP; 1774 A; 1494 C; 1784 G; 1365 T; 0 U; 0 Other;
SQ	Query Match 10.1%; Score 421.4; DB 6; Length 6417; Best Local Similarity 51.9%; Pred. No. 1.4e-101; Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;
QY	818 TACATCCATACACAGCTTGAGGAGCACTTAATTTTTTTCGGGTTCTTCGTGGTCAAAACAGACGC 877
Db	2266 TGCACCCCTATCAATGAGGGCGCTGAATGGTTTGGCGTTCTCTCGGGCTCAGGCGCACTG 2325
QY	878 ATGTAATCCTTGCTGATGAATGGGACTGAGCAAGACAAATCAAGCAATGCCCTTTTAG 937
Db	2326 ACACCAATTTTGGCTGATGAGATGGGCTTTGGGAAACTGTACAGACAGCACTTCTCTGTT 2385

Db 3454 TCCTGCTTCCACTCGAGCTGGGGCCCTTGAATCAATCTGGCCACTGCTGACACAGTTA 3513
Qy 2060 TCATTATGACAGTGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
Db 3514 TTATCTATGACTCTGACTGGAACCCCAATATGACATTCAGGCCCTTTAGCAGAGCTCAC 3573
Qy 2120 GACTTGGCCCAACAAATAAGTGTATGATTTATAGGCTCATAAACCGAGGCACCAATTCGAAG 2179
Db 3574 GGATTGGGCAAAATAAAGAGTAATGATCTACCGGTTTGACCCGTCGTCAGTGGAGG 3633
Qy 2180 AAAGGATGATGCAATTGACTAAAGAAATATGGTTCTAGAGCATCTTGTGTTGG----- 2234
Db 3634 AGCGCATCACGAGGTGGCAAGAGAAATATGATGCTGACGCACTCTAGTGTGGCGGCTG 3693
Qy 2235 -----GAAACTCAAAACACAAACATTAATCAGGAAGATTAGATGACATCATCGGTATG 2290
Db 3694 GGCTGGGCTCAAGACTGGATCTATGTCCAAACAGGAGCTTGATGATATCCTCAAAATTG 3753
Qy 2291 GATCAAGAGGAGCTTTTGTAGTGAAGATGATGAAGCAGAAAGTCTGGAAAAATTCATT 2350
Db 3753 GCATGAGGAACATTTCAAGGATGAAGCCACTGATGAGGAGGAGACAAACAGAGGGAG 3813
Qy 2351 ATGATGATGGCGCTATAGACAAATTTGCTTGA-----TCGTGATCTCGTGGAGGCAGG 2404
Db 3814 AAGATGACAGTGTATCCACTACGATGAAGGCCATTGAAGCGCTGTAGACCGTAACC 3873
Qy 2405 AAGTCTCAGTGGATGATGAGAGAGAGATGATTTCTTAAAGGCTTTCAAGGTGCTTAATT 2464
Db 3874 AGGATGAGACTGAAGACACAGAAATTCAGGCGCATGAATATTTGAGCTCAATTCAAAG 3933
Qy 2465 TTGAATATATAGTGAATGAGGAGCAGCAGCATTTAGAGGCACAGAGATCGCTCTGAAA 2524
Db 3934 TGGCCCATGATGTGTCAGGGAAGAAATGCGGGAGGAGAGGATGAACCGGAAA 3993
Qy 2525 GCAAAATCTTCAGCAGGCAATTCGTATAGAGCAAGTTATTTGGGAAGAGTTGTTTAAAGATA 2584
Db 3994 TCATTAAACAGGAAGAAGTGTGATCCTGA---CTACTGGAGAAATTCGTCGGCACC 4050
Qy 2585 AATTGAGTGCACAGGCTGAGAGCTTAATGCTCTTGGAAAGAGAGAGATCGCA 2644
Db 4051 ATTATGAGCAGCAGCAAGAGATAGCCCGAAATCTGGCAAGAGAAAGATCCGTA 4110
Qy 2645 AGCAGTGGTATCATTTGAAG--AAGATGATCTTGTGTTGGAAGATGTAGCTG 2701
Db 4111 AACAGGCTCACTAATGATGGCTCCAGGAGGACCGAGATTGCGAGCAGCACAGTCCG 4170
Qy 2702 ATGGAGATGAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAGGAGTTC 2761
Db 4171 ACAACAGTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTTGATGAACGTT 4230
Qy 2762 AGACGGTGCAGCGCGGTACAGAAAGAGGTGCGGATAATTTGGAAACCAA----- 2812
Db 4231 CAGAAGTCTCCCGTAGGCCAGCTGTAAGGCCCTGCGGAATGATAAAGATAAGACCATTCG 4290
Qy 2813 CTCGCTGATGAAGGTGAGGGAGATCTTTACAGTACTGGGTTTCAACACAGAGTCAAA 2872
Db 4291 CTCCTCTGTGGCCGCTGTTGGTGGGAATATGAAGTACTTGGTTTAAATGCTCGTCAGC 4350
Qy 2873 GGGCCATTTTGTACAGACTTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT 2923
Db 4351 GAAAGGCTTTCTTAATGCAATTTATGCAATGATGATGCCACCTCAGGATGCTTTTACTA 4410
Qy 2924 GGAAGGAGTTTGTCTCGCTTAAAGCAGAGAGCTTTGAAGAAATTAATGAATATGAA 2983
Db 4411 CCCAGTGGCTTGTGAAGAGACCTGGGAGCAATCAGAGAAAGAGTTCAAGGCATATGCT 4470
Qy 2984 TACTCTTCTCAACACATTCCTGAAGAAATAGACAGAGAAATTCCTCAACCTTTTCAGATG 3043
Db 4471 CTCCTTTTCATCGGCAATTTATGAGCGCGGGGAGATGGGGCTTGAGACCTTTTCTGATG 4530
Qy 3044 GTGTGCCCAAGGAAGGACTTGAATAGAAAGATGTTCTAGTCAGAAATGCTCTTCTGATAC 3103

Db 4531 GTGTCCCGCAGAGGCGCTGTCTCGCCAGCATGCTCTACTAGATTGGTGTATGCTT 4590
Qy 3104 TAGTTCCAGGAGAGGTGAATTTGTAGAAGATCATCCAGGG 3144
Db 4591 TGATTCCGAAGAGGTTCCAGGAGTTGAACATGTTAATGGG 4631
RESULT 7
ABK84696
ID ABK84696 standard; cDNA; 6417 BP.
XX ABK84696;
AC AC
XX AC
DT 14-AUG-2002 (first entry)
XX Human cDNA differentially expressed in granulocytic cells #1267.
DE Human; ss; granulocytic cell; DNA chip; bacterial infection;
XX viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac respiratory injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX Homo sapiens.
OS
XX
XX WO200228999-A2.
PN 11-APR-2002.
XX
XX 03-OCT-2001; 2001WO-US030821.
PF
XX 03-OCT-2000; 2000US-0237189P.
PR
XX (GENE-) GENE LOGIC INC.
XX
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
PI WPI; 2002-435328/46.
XX
XX Detecting granulocyte activation by detecting differential expression of genes associated with granulocyte activation, which serves as diagnostic markers that is useful for monitoring disease states and drug toxicity.
PS Claim 1; SEQ ID NO 1267; 114pp; English.
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation (GCA), by detecting the level of expression of gene(s) (Gs) identified by DNA chip analysis as given in the specification, and comparing the expression level to an expression level in an unactivated GC, where differential expression of Gs is indicative of GCA. Also included are modulating (M2) GA by contacting GC with an agent that alters the expression of at least one gene in Gs; (2) screening (M3) for an agent capable of modulating GCA or an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease using the gene expression profile; (3) detecting (M4) an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by detecting the level of expression in a sample of the tissue of gene(s) from Gs, where the level of expression of the gene is indicative of inflammation; (4) treating (M5) an inflammation (especially chronic) or in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by contacting a tissue having inflammation with an agent that modulates the expression of gene(s) from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for modulating GCA; M3 is useful for screening an agent capable of modulating GCA preferably in an inflammation in a tissue; M4 is useful for detecting an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease (e.g. psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,

CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC disease, ulcerative colitis, periodontal disease; also bacterial
CC infection, viral infection, parasitic infection, protozoal infection,
CC fungal infection and M5 is useful for treating one of the above
CC conditions. The present sequence represents a gene differentially
CC expressed in granulocytes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6417 BP; 1774 A; 1494 C; 1784 G; 1365 T; 0 U; 0 Other;
Query Match 10.1%; Score 421.4; DB 6; Length 6417;
Best Local Similarity 51.9%; Pred. No. 1.4e-101;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;
818 TACATCCATACCGCTGAGGAGCTTAATTTTGGCGTCTCGTGGTCAAAACAGCGC 877
2266 TGCACCCCTATCAAAATGGAGGCGCTGAATTTGGTTCGCGCTTCTCTCGGCTCAGGCACTG 2325
878 ATGTAATCCCTTGTGTGTAATGGAATAGGCACTAGGCAAGCAATTCAAAGCATTCGCCCTTTTAG 937
2326 ACACATCTTGGCTGATGAGTGGCCCTTGGAAACTGTACAGACAGCAGTCTTCCTGT 2385
938 CTTACATTTTGGAGGAG-----AACCTCATTCGCGAATTTGGTAATTTGCTCCTCATCGA 991
2386 ATTCCCTTTCAAGAGGGGTCAATTCCAAAGGCCCTTCTAGTGAGGCGCCCTCTTTCTA 2445
992 CTCTGCGTAAGTGGAGAGAGTTTCCCATGTCAGAGTATTCAGAGAACTGATTTTACTTATCGAAAG 1111
2446 CCATCATCAATGGGAGCGGAGTTTGAATGTGGGCTCCAGACATGATGTGCTAACCT 2505
1052 ATTTTGCACCTGGCCAAAGCTGAGCAGTATTCAGAGAACTGATTTTACTTATCGAAAG 1111
2506 ATGTGGGTGCAAGGACAGCGGTGCCATCATCCGAGAGATGATTTCTCTTTGAAGACA 2565
1112 ATCAAAAAAAGATCAAGAAAAAAGAAATCTGGACAAATTAAGTAGCGAAAGCAAGCAAAAA 1171
2566 ATGCCATTCGTGGTGGCAAGAGGCTC-----CCGCATGAAGAAAGAGGCAT 2613
1172 GAATCAAGTTTGTGTCCTCTCATATCGTATGATGATGATCAACCTAGATTCAGCAGTTC 1231
2614 CTGTGAATTTCCATTTGCTGCTGATCATCTTATGAATTTGATCACCATTGATGCTGCTATTT 2673
1232 TAAAAACCAATTAAGTGGGAGTGCATGATTTGTGATGAAGGTTCATCGACTGAAAAATAAG 1291
2674 TGGCTCTATTTGATTTGGGCTGCTCATCTGTTGATGAAGCCCATCGGCTGAGNACATC 2733
1292 ATTCAAAGCTGTTCTTCTCATTTGACACAGTATTCAGTAATCAACCGCTATTCTTCTGACAG 1351
2734 AGTCTAAGTTCTTCGGGTATTGAATGTTACTCACTCCAGCACAAAGCTGTTGCTGACTG 2793
1352 GAACACACTTCAGAACAACTTTGATGAACTTTTTCATGCTCATGCAATTTTCTTGATCGG 1411
2794 GGACACCATTTACAAAAAATCTGGAAGAGTTGTTTCTATCTGCTCAACTTTCTCACCCCGC 2853
1412 GGAAGTTTGGAGTTTGGAGAGTTCCAGGAGGAGTTTCAAGATATTAAATCAAGAGAGC 1471
2854 AGAGTTTCCAAATTTGGAAGTTTGGAGGAGTTTGGAGGAGTTTGTGACATTTGCCNAGAGAGC 2913
1472 AGATCTCAAGGTTGCACAAATTTTGGCTCCACATTTTGGCTCAGAGGAGTTAAAAAAGACG 1531
2914 AGATAAAAAAATTCATGACATGCTGGGGCGGCATGTTTGGCGCGCTCAAGCCGATG 2973
1532 TAATGAAGACATGCCCCCAAAAGAGCTCATTTTGGCGTTGATGTGAGCAGTCTGC 1591
2974 TGTTCAGAAATGCTCCCTTCCAAGACAGAACTAAATTTGTGCGTGTGGAGCTGAGCCTATGC 3033
1592 AGAAGATATTAACAGCTATTTTACCCTGAATTTATCAAGTATTCG---CAAAAAAGG 1648
3034 AGAAGAAATACTACAAGTACATCTCTCACTCGAAATTTTGAAGCACTCAATGCTCCCGAGGTG 3093

QY 1649 GAGGTGCTCAAAATTTCCCTTAATAACATTTATGATGGAATACGAAAGTATGCTGCCATC 1708
DB 3094 GTGGCAACAGGTGCTCTGCTGAATGTGGTGAATGATCTTAAGAAAGTGTCTGCAACATC 3153
QY 1709 CTTA-----TATGCTAGAGGGTGTGGAGCCAGTTATTCACGACGCAAAATGAAGCTT 1759
DB 3154 CATACCTCTTCCCTGTGGCTGCATGGAAGCTCTTAAGATGCTTAATGGCATGTATGATG 3213
QY 1760 TCAAAACAATTTTGGAGTCTTGTGGAAAGCTGCAACTTCTAGATAAAATGATGGTCAAAC 1819
DB 3214 GCAGTCCCTAATCAGAGCATCTGGGAAATATTGCTGTCGAGAAAAATGCTCAAGAAC 3273
QY 1820 TGAAGAGCAAGGACACAGAGTCTTAATATACACAGTTTCAAGCATATGCTGCACTTAC 1879
DB 3274 TTAAGAGAGGTGGGCATCTGTACTCATCTTTCCAGATGACCAAGATGCTAGACCTGC 3333
QY 1880 TTGAAGACTACTGTACCCATAGAAATGCGAGTACGAGGGAATTTGATGGAAGAGTTGGCG 1939
DB 3334 TAGAGGATTTCTTGGAAACATGAAGTTTATAATACGACGATCGATGGTGAATCACTG 3393
QY 1940 GAGCTGAGCGGCAAAATACGATAGATCGGTTCAATGCCAAAAATTTCTAACAGTTTGT 1999
DB 3394 GGAACATGCGGCAAGAGGCAATTCACCGCTTCAATGCAACCGGTGCTCAGCAGTTCTGCT 3453
QY 2000 TTTTGTCTCCCAAGAGCTGTGGCTTAGGATTAATCTTCAACGCGCTGATACAGTAA 2059
DB 3454 TCTTGTCTTCCATCGAGCTGGGGCTTTGGAATCAATCTGGCCACTGCTGACACAGTTA 3513
QY 2060 TCATTTATGACAGTGAAGTCCATCTGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
DB 3514 TTATCTATGACTGACTGGAACCCCATTAATGACATTCAGGCTTTAGCAGAGCTCACC 3573
QY 2120 GACTTGGCCAAAACAAATAAGTGAATTTATAGGCTCTATAAGCTCTATAACCGAGGCCATTCATTGAAG 2179
DB 3574 GGATTGGCAAAATAAAAGGTAATGATCTACCGGTTTGTGACCGCTGCTCAGTGAGG 3633
QY 2180 AAAGGATGATGCAATTTGACTATAAAAGAAATGTTCTTAGAGCATCTTGTGTTGG----- 2234
DB 3634 AGCGCATACGCGAGTGGCAAGAAATGATGCTGACGCACTAGTGTGTCGCGCTG 3693
QY 2235 ----GAAACTCAAAACACAAAAACATTAATCAGGAAGTTTAGATGACATCATCAGGTATG 2290
DB 3694 GGTGGGCTCCAAGACTGGATCTATGTCCAACAGAGCTTTGATGATATCTCAATTTG 3753
QY 2291 GATCAAGAGGAGCTTTTGTGTGTAAGATGATGAAGCAGGAAAGTCTGGAAAAATTCATT 2350
DB 3754 GCACCTGAGGAATTTTCAAGGATGAAGCCACTGATGGAGGAGGAGACAAACAAAGAGGAG 3813
QY 2351 ATGATGATGCGGCTATAGACAAATTTGCTTGA-----TCGTGATCTCGTGGAGGCAGAG 2404
DB 3814 AAGATAGCAGTGTATTCCACTACGATGATAAGGCCATTGAACCGCTGCTAGACCGTAACC 3873
QY 2405 AAGTCTCAGTGGATGATGAAGAGGAGATGATTTCTTAAAGGCTTTCAAGGTGGCTAAT 2464
DB 3874 AGGATGAGACTGAAGACACAGAAATTCAGGGCAATGATGATTAATTTGAGCTCATTCAGAG 3933
QY 2465 TTCAATATATAGATGAAAAATGAGCAGCAGCATTAGAGGCACAGAGAGTCTGCTGTA 2524
DB 3934 TGSCCCCATGATGTGGTACGGGAAGAAATGGGGGAGGAAGAGGAGTAGAAGCGGAAA 3993
QY 2525 GCAATCTTCAGCAGGCAATTTCTGATAGAGCAAGTTATTGGGAAGAGTTGTTTAAAGATA 2584
DB 3994 TCATTTAAACAGGAAGAAAGTGTGGATCCCTGA---CTACTGGGAGAAATTTGCTCGGCACC 4050
QY 2585 AATTTGAGCTGCACCGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGTACGCA 2644
DB 4051 ATTATGAGCAGCAGCAGAAAGATCTAGCCCGAAATCTGGCAAAAGGAAGAAATTCGTA 4110
QY 2645 AGCAGTTGGTATCCATTTGAAG---AAGATGATCTTGTGTTTGGAGATGTGAGCTCTG 2701
DB 4111 AACAGGTCAACTACATGATGGCTCCAGGAGGAGCCGAGATTCGAGGACGACAGTCCG 4170
QY 2702 ATGGAGATGAAGTTTATGAAGCTGATGATCAACAGATGTTGAAGCAGCAGGAGGAGGATTC 2761

Db 4171 ACAACAGTCCGATTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTTGATGAACGTT 4230
Qy 2762 AGACGGGTGACGCGCGGTACAGAAGAAGGTCCGGATAATTTGGAAACCA----- 2812
Db 4231 CAGAAGTCCCGGTAGGCCCAAGTCGTAAAGGCCCTGCGGAATGATAAAGATAAGCCATTGC 4290
Qy 2813 CTCGGTTGATGGAAGGTGAGGGAGATCTTTTCAGACTACTGGGTTTCAACCCAGAGTCAAA 2872
Db 4291 CTCCTCTGTGGCCGGTGTGGTGGGAATATTGAAGTACTTGGTTTAAATGCTCGTCAGC 4350
Qy 2873 GGGCCATTTTGTACAGACTTTGTAGAGGTATGG-----AGCTGGCAATTTTGATT 2923
Db 4351 GAAAAGCCTTCTTAATGCAATTAATGATATGATGTCACACCTCAGGATGCTTTTACTA 4410
Qy 2924 GGAAGGAGTTTGTCTCGCTTAAAGCAGAGACCTTTGAAGAAATAAATGAATATGAA 2983
Db 4411 CCCAGTGGCTTGTAAAGACCTGCGAGCAATCAGAGAAGAGTTTCAAGGCATATGTCT 4470
Qy 2984 TACTCTTCTTGAAGCACATTGCTGAAGAAATAGACGAGAATTTCTCCAACTTTTTCAGATG 3043
Db 4471 CTCCTTTTCAATGCGATTTATGTAGCGGGGCGAGATGGGCGTGAGACCTTTTGTCTGATG 4530
Qy 3044 GTGTGCCCAAGGAAGGACTTAGAATAGAAAGATGTTCTTAGTCAGAAATGCTCTTCTGTATAC 3103
Db 4531 GTGTCCCGCGAAGGCGCTGTCTCGCCAGCATGTCCTTACTAGAAATTTGGTGTATGTCTT 4590
Qy 3104 TAGTTCAGGAAGAGTGAATTTGTAGAAGATCATCCAGGG 3144
Db 4591 TGATTGCAAGAAGGTTTCAGGAGTTTGAACATGTTAATGGG 4631

RESULT 8

AAD54631 standard; DNA; 6417 BP.
XX AAD54631;
XX 26-JUN-2003 (first entry)
XX Human chromodomain helicase DNA binding protein (CHD) encoding DNA #6.
DE Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;
KW chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;
KW cell proliferative disorder; chromatin organisation modifier domain;
KW cytoskeletal; ds.
XX Homo sapiens.
OS WO200298899-A2.
XX 12-DEC-2002.
XX 03-JUN-2002; 2002WO-US017466.
XX 05-JUN-2001; 2001US-0296076P.
XX 10-OCT-2001; 2001US-0328605P.
XX 22-OCT-2001; 2001US-0338733P.
XX 15-FEB-2002; 2002US-0357253P.
XX 15-FEB-2002; 2002US-0357600P.
XX (EXEL-) EXELIXIS INC.
XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
PI Lioubin MN;
XX WPI; 2003-156840/15.
XX Identifying a candidate p53 pathway-modulating agent as therapeutic
PT targets for disorders related to defective p53 function e.g. cancer by
PT contacting an assay system having purified CHD polypeptide or nucleic
PT acid, with a test agent.

PS Disclosure; Page 59-62; 278pp; English.
XX The present invention relates to a method for identifying candidate p53
CC pathway modulating agents. The method involves contacting an assay system
CC comprising purified chromatin organisation modifier (chromo) domain
CC helicase DNA binding proteins (CHD), nucleic acids, their functionally
CC active fragments or derivatives, with a test agent under conditions
CC where, but for the presence of the test agent, the system provides a
CC reference activity. The methods are useful for identifying modulators of
CC the p53 pathway as therapeutic targets for disorders associated with
CC defective p53 function, such as angiogenic disorders, apoptotic disorders
CC or cell proliferative disorders, e.g. cancer. The modulators are useful
CC as research reagents, diagnostics and therapeutics. The invention is also
CC useful in gene therapy. The present sequence is human CHD DNA
XX
SQ Sequence 6417 BP; 1774 A; 1494 C; 1784 G; 1365 T; 0 U; 0 Other;
Query Match 10.1%; Score 421.4; DB 7; Length 6417;
Best Local Similarity 51.9%; Pred. No. 1.4e-101;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;
Qy 818 TACATCATACAGCTTGAGGGACTTAATTTTGGGGTCTCGTGGTCAAAACACAGCG 877
Db 2266 TGCACCCCTATCAATGAGGGCTGAATTTGGTGGGCTTCTCTGGGCTCAGGCACTG 2325
Qy 878 ATGTAATCTTGTCTGATGAAATGGGACTAGGCAAGACAATTCAAAGCATTTGCCCTTTTAG 937
Db 2326 ACACCATCTTGGCTGATGATGGGCTTTGGGAAAACCTGTACACAGACAGTCTTCTGT 2385
Qy 938 CTTTCACTTTTGTAGGAG-----AACCTCATTCGCAATTTGGTAAATTTGCTCTATPCA 991
Db 2386 ATTCCCTTTTACAAGGAGGGTCAATTCAAAGGCCCTTCTCTAGTGAGCGCCCTTTCTA 2445
Qy 992 CTCGTGTAACCTGGAGAGAGATTTGCCACATGCGGCCCCACAGATGAAGCTGTTATGT 1051
Db 2446 CCATCATCACTGGGAGCGGGAGTTTGAATTTGGGCTCCAGACATGTATGTCTAACCT 2505
Qy 1052 ATTTTGGCACTGCGCAAGCTTCGAGCAGTTATCAGAGAACTAGTGTATTTACTTATCGAAAG 1111
Db 2506 ATGTGGTGCACAGGACAGCGGTGCCATCATCCGAGAGATGATGTTCTCTTTGAAGACA 2565
Qy 1112 ATCAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAGCAAAAA 1171
Db 2566 ATGCCATTCGTGTGGCAAGAGGCCCTC-----CCGCATGAAGAAAGAGCAT 2613
Qy 1172 GAATCAAGTTTGTATGCTCTCTCACATCGTATGAGATGATCAACCTAGATTTCAGCAGTTC 1231
Db 2614 CTGTGAATTTCCATGTGCTGACATCTTATGAATTTGATCACCATTGACATGGCTATTT 2673
Qy 1232 TAAAAACCAATTAAGTGGAGTGCATGATTTGTCATGAAGTTCATCGACTGAAAAATAAGG 1291
Db 2674 TGGGCTCTATTGATTGGGCTGCTCATCGTGATGAGCCCATCGGTGAGAACATC 2733
Qy 1292 ATTCAAGCTGTTCTCTTCATTGACACAGTATTCAAGTAACCAACCGTATTTCTTGACAG 1351
Db 2734 AGTCTAAGTTCTTCCGGTATTGAATGGTTACTCACTCCAGCAACAAGCTGTTGCTGACTG 2793
Qy 1352 GAACACCACTTCAGAACCACTTGGATGAACCTTTTATGTCATGCACTATTTCTTGATCGCG 1411
Db 2794 GGACACCATTTACAAAAAATCTGGAAGAGTTGTTTTCATCTGCTCAACTTTCTCACCCCG 2853
Qy 1412 GGAAGTTTGGAGTTTGGAGGAGTTCCAGGAGAGTTTCAAGATATTTAATCAAGAGAGC 1471
Db 2854 AGAGTTCCCAATTTTGGAAAGTTTGGAGAGAGTTTGTGACATTTGCCAGAGAGACC 2913
Qy 1472 AGATCTCAAGGTTGCACAAAAATGTTGGCTCCACATTTGCTCAGAGAGGTTAAAAAGAGC 1531
Db 2914 AGATAAAAAAAGTGCATGACATGCTGGGCGGCACATGTTGCGGCGGCTCAAAAGCCGATG 2973
Qy 1532 TAATGAAGACATGCCCCCAAGAGAGCTCAATTTTGGTGTGTTGATCTGACAGCAGTCTGC 1591
Db 2974 TGTTCAAGAACATGCCCCCTCCAGACAGAACTAATTTGTGGGTGGAGCTGAGCCCTATGC 3033

PT targets for disorders related to defective p53 function e.g. cancer by
PT contacting an assay system having purified CHD polypeptide or nucleic
PT acid, with a test agent.

XX Disclosure; Page 62-66; 278pp; English.

PS The present invention relates to a method for identifying candidate p53
CC pathway modulating agents. The method involves contacting an assay system
CC comprising purified chromatin organisation modifier (chromo) domain
CC helixase DNA binding proteins (CHD), nucleic acids, their functionally
CC active fragments or derivatives, with a test agent under conditions
CC where, but for the presence of the test agent, the system provides a
CC reference activity. The methods are useful for identifying modulators of
CC the p53 pathway as therapeutic targets for disorders associated with
CC defective p53 function, such as angiogenic disorders, apoptotic disorders
CC or cell proliferative disorders, e.g. cancer. The modulators are useful
CC as research reagents, diagnostics and therapeutics. The invention is also
CC useful in gene therapy. The present sequence is human CHD DNA

XX Sequence 6417 BP; 1774 A; 1494 C; 1784 G; 1365 T; 0 U; 0 Other;

Query Match 10.1%; Score 421.4; DB 7; Length 6417;
Best Local Similarity 51.9%; Pred. No. 1.4e-101;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;

Qy	818	TACATCCATCCAGCTTGAGGAGCTTAATTTTGGCGTCTCGTGGTCAAAACAGACGC	877
Db	2266	TGCACCCCTAATCAATGAGGCGCTGAATGGTTGGCTCTCTCGGCTCAGGCGACTG	2325
Qy	878	ATGTAATCCTTGGTGAATGGGACTAGGCAAGACAAATTCAGAGCATTTGCCCTTTAG	937
Db	2326	ACACCACTTGGCTGATGAGTGGCCTTGGGAACTGTACACAGCAGCTTTCCTGT	2385
Qy	938	CTTCACCTTTTGGAGAG-----AACCTCATTCGCGCATTTGGTAAATTCCTCTATCGA	991
Db	2386	ATTCCCTTTTACAGAGGAGGTCAITCCAAAGGCCCTTCTAGTGGCGGCCCTCTTTCTA	2445
Qy	992	CTCTGCGTAACTGGGAGAGAGTTGCCATATCCAGATGCGCCCAACATGAATGGTATGT	1051
Db	2446	CCATCATCACTGGGAGCGGAGTTGAATGTGGGCTCCAGACATGATGTGCTAACT	2505
Qy	1052	ATTTTGGCAGCTGCGCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTTACTTATCGAAAG	1111
Db	2506	ATGTGGGTGCAAGGACAGCGCTGCCATATCCGAGAGATGAGTTCTCTTTTGAAGACA	2565
Qy	1112	ATCAAAAAGACATCAAGAAAAGAAATCTGCACAAATAAGTAGCGAAGCAGCAAAAAA	1171
Db	2566	ATGCCATTCGTGGTGGCAAGAGGCCTC-----CCGCATGAAGAAAGAGGCAT	2613
Qy	1172	GAATCAAGTTTGTCTCTCTCATGATGATGATGATCAACCTAGATTCAGCAGTTC	1231
Db	2614	CTGTGAATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2673
Qy	1232	TAAACCAATTAAGTGGGAGTGCATGATTTGTGATGAAGGTTCATCGACTGAAATAAGG	1291
Db	2674	TGGGCTCTATTGATTGGGCTGCTCATCGTGGATGAAGCCCATCGGCTGAAGAACATC	2733
Qy	1292	ATTCAGAGCTTCTCTTCATTTGACACAGTATTCAAGTAACCCGATTTCTTCTGACAG	1351
Db	2734	AGTCTAAGTTCTTCGGGTATTGAAATGGTTACTCACTCCAGCAAAAGCTGTGCTGACTG	2793
Qy	1352	GAACACCACTTTCAGAACCACTTGGATGAATTTTTCATGCTCATGATTTCTTCATGCGG	1411
Db	2794	GGACACCACTTTCAGAACCACTTGGATGAATTTTTCATGCTCATGATTTCTTCATGCGG	2853
Qy	1412	GGAAATTTGGAGTTTGGAGAGTTTCCAGGAGGAGTTCAAGATATTAAATCAAGAGAGC	1471
Db	2854	AGAGTTTCCACAAATTTGGAAGTTTGGAGGAGTTTGGTGTGCTGACATTTGCCAAGAGGACC	2913
Qy	1472	AGATCTCAAGGTTGCACAAATTTGGCTCCACATTTGCTCAGAGGTTAAAAAGACG	1531
Db	2914	AGATAAAAAAATTCGATGACATGCTGGGGCCGACATGTTGGCGGCTCAAGCCGATG	2973

Qy	1532	TAATGAAGACATGCCCCCAAAAAAGAGCTCATTTTGGTGTGTGATCTGAGCAGTCTGC	1591
Db	2974	TGTTCAAGAACATGCCCCCTCAAGACAGAACTAATTTGTCGTGTGGAGCTGAGCCCTATGC	3033
Qy	1592	AGAAAGAAATTTACAAAGCTATTTTACCCGTTAATTTATCAAGTATTGA---CAAAAAGG	1648
Db	3034	AGAAGAAATACTACAAAGTACATCTCACTCGAAATTTTGAAGCACTCAATGCCGAGGTG	3093
Qy	1649	GAGGTCTCATAATTTCCCTTAATAACATTTATGATGAAATTTACGAAAGTATGTCGCCATC	1708
Db	3094	GTGGCAACAGGTTCTCTGCTGAATTTGTTGATGATGATCTTAAGAAGTGTCTGCAACATC	3153
Qy	1709	CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTCACGACGCAAAATGAAGCTT	1759
Db	3154	CATACCTTTCCTCTGGCTGCAATGGAAGCTCTTAAGATGCCCTAATGGCATGTATGATG	3213
Qy	1760	TCAACAACCTTTTGGAGTCTTTGGAAAGCTGCAACTCTTAGATAAATATGATGGTCAAC	1819
Db	3214	GCAGTCCCTTAATCAGAGCATCTGGGAAATTTATTTGCTGCTGCAGAAATTTGCTCAAGAACC	3273
Qy	1820	TGAAGAGCAAGGACACAGAGTCTTAATATACACAGTTTTCAGCATATGCTGCGACTTAC	1879
Db	3274	TTAAGAGAGGTGGGCATCGTGTACTCATCTTTCCAGATGACCAAGATGCTAGACTGC	3333
Qy	1880	TTCAAGACTACTGTACCCCATTAAGAAATGGCAGTACGAGCGAATTTGATGAAAGTTGGCG	1939
Db	3334	TAGAGGATTTCTTGGAAACATGAAGTTATAAATACGAACGATCGATGGTGAATCACTG	3393
Qy	1940	GAGCTGAGCGGCAAAATACGATATGATCGGTTCAATGCCAAAAATTTCTAACAGTTTGT	1999
Db	3394	GGAAATCTGGCGCAAGAGGCCATTTGACCGCTTCAATGACCGGCTGCTCAGCAGTTCTGCT	3453
Qy	2000	TTTTGCTCTCCACAGAGCTGGTGGCTTAGGAAATAAATTTTCAACCGCTGATACAGTAA	2059
Db	3454	TCCTGCTTTCCATCGAGCTGGGGGCTTTGGAATCAATCTGGCCACTGCTGACACAGTTA	3513
Qy	2060	TCAATTTATGACAGTGTGAATTCCTCATGCTCATCTTCAAGCAATGGCTAGAGCTCATC	2119
Db	3514	TTATCTATGACTCTGACTGGAAACCCCAATAATGACATTCAGGCCCTTTAGCAGAGCTCAC	3573
Qy	2120	GACTTGGCCMAAATAAGTGTATGATTTATAGGCTCATAAACCGAGGCCATTTGAAG	2179
Db	3574	GGATTGGGCAAAATAAAAGGTAAATGATCTACCGTTTGTGACCGCTGCTCAGTGAGG	3633
Qy	2180	AAAGGATGATGCAATTTGACTTAAAGAAATGTTCTAGAGCATCTTGTGTTGG-----	2234
Db	3634	AGCGCATCACGAGGTGGCMAAGAAATAATGATGCTGACGCTATGCTGCTGGCGCTG	3693
Qy	2235	---GAAACTCAAAACACAAACATTAATCAGGAAGATTAGATGACATCATCAGGTATG	2290
Db	3694	GGCTGGGCTCCAAGACTGGATCTATGTCCAACACAGGAGCTTGATGATATCTCAAAATTTG	3753
Qy	2291	GATCAAGAGGCTTTTGTGTAGTGAAGATGATCAAGCAGGAAAGTCTGGAAAAATTCATT	2350
Db	3754	GCATCGAGGAATTAATCAAGGATGAAGCCACTGATGGAGGAGGAGCAACAAAGAGGAG	3813
Qy	2351	ATGATGATGGCTATAGACAAATTTGCTGA-----TCGTGATCTCGTGGAGCAGAGG	2404
Db	3814	AGATAGCAGTGTATTCACCTACATGATGAAGCCATTGAACGGCTGTAGACCGTAACC	3873
Qy	2405	AAGTCTCAGTGGATGAAGAGGAGATGGATTTCTTAAAGGCTTTCAAGGTGGCTTAAT	2464
Db	3874	AGGATGAGACTGAAGACACAGAAATTCAGGGGCATGAATGAATATTTGAGCTCATTTCAAAG	3933
Qy	2465	TTGAATATATAGTAAATGAGCAGCAGCATTTAGGGCAGCAGAGATGCTGCTGAAA	2524
Db	3934	TGGCCCCAGTATGTGTCACGGNAAGAAATGGGGGAGGAGGAGGTAGAACGGGAAA	3993
Qy	2525	GCAAACTTTCAGCAGGCAATTTCTGATAGACAAAGTTATTTGGGAAGAGTTTGAAGATA	2584
Db	3994	TCATTAACAGGAAAGAAAGTGTGATCTCTGA-----CTTACTGGAGAAATTTGCTGGGACC	4050
Qy	2585	AATTTGAGCTGCACCGGCTGAGGAGCTTAATGCTCTTTGGAAAAAGGAGAGAGTCGCA	2644

```
Db 4051 ATATGAGCAGCAGCAGAGATCTAGCCGAATCTGGGCAAGAGAAAGATCCGTA 4110
Qy 2645 AGCATGTTGGTATCAATGAAG---AAGATGATCTGCTGTTGGTGAAGATGTGAGCTCTG 2701
Db 4111 AACAGGTCAACTCAATGATGCTCCAGGAGGACCGAGATTGGCAGGACGACCAAGTCCG 4170
Qy 2702 ATGAGATGAAGATTGAGCTGAGTCAACAGATGTTGAAGCAGCAGGACAGGAGTTC 2761
Db 4171 ACAACAGTCCGATTACTCATGCTGCTTCAGAGGAAGGTGATGAAGACTTTTGATGAACGTT 4230
Qy 2762 AGACGGTTCAGCGCCGCTACAGAGAAAGGTCGCGATAATTGGAAACCAA----- 2812
Db 4231 CAGAGCTCCCGTAGGCCAGTCTGAAGGCCCTGCGGATGATAAAGATGAAGCATTGC 4290
Qy 2813 CTCGTTGATGGAAGGTGAGGGAGATCTTTTCAGAGTACTGGGTTTCAACCAGAGTCAAA 2872
Db 4291 CTCCTCTGTTGGCCGCTGTTGGTGGGAATATGAAGTACTTGGTTTTAATGCTCGTCAGC 4350
Qy 2873 GGGCCATTTTGTACAGACTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT 2923
Db 4351 GAAAGAGCTTTCTTAATGCAATATATGCGATATGATGCCACCTCAGGATGCTTTTACTA 4410
Qy 2924 GGAAGGAGTTTGTCTCTGCTTAAAGCAGAGACCTTTTGAAGAAATAAATGAATGGA 2983
Db 4411 CCCAGTGGCTTGTGAAGAGACCTGCGAGGCAATCAGAGAAGATTCAGGCAATATGCT 4470
Qy 2984 TACTCTTCTTGAACACATGCTCAGAAATATAGACGAGAATCTTCAACCTTTTCAGATG 3043
Db 4471 CTCCTTTTCATGCGCAATTTATGTAGCGCGGGCAGATGGGCTGAGACCTTTGCTGATG 4530
Qy 3044 GTGTGCCCCAAGGAGGACTTAGAATAGAAGATGTTCTTAGTCAGAAATGCTTCTGTATAC 3103
Db 4531 GTGTGCCCCGAGAAGGCTGTCTGCCAGCATGTCTTACTAGAAATTTGGTTATGTCTT 4590
Qy 3104 TAGTTCAGGAGAGGTGAAATTTGAGAAGATCATCCAGG 3144
Db 4591 TGATTGCAAGAGGTTTCAGGAGTTTGAACATGTTAATGGG 4631

RESULT 10
AA158444
ID AA158444 standard; cdna; 6475 BP.
XX
AC AA158444;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 647.
XX
KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
```

```
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM39288.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
PS Claim 1; SEQ ID NO 647; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nontropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 6475 BP; 1760 A; 1528 C; 1822 G; 1365 T; 0 U; 0 Other;
Query Match 10.1%; Score 421.4; DB 4; Length 6475;
Best Local Similarity 51.9%; Pred. No. 1.4e-101; Indels 69; Gaps 10;
Matches 1236; Conservative 0; Mismatches 1076;
Qy 818 TACATCCATACCCAGCTTGAGGGACTTAATTTTTCGGTTCCTCGTCAAAACAGACGC 877
Db 2395 TGCACCCCTATCAATGGAGGGCTCGAATTTGGTTGGCTTCTCTGGGCTCAGGCGACTG 2454
Qy 878 ATGTAATCCTTGCTGATGAATGGGACTAGGCAAGACAATTCAGAGCATTCGCTTTTAG 937
Db 2455 ACACCATCTTGGCTGATGAGATGGGCTTGGGAAACTGTACAGACAGCAGTCTTCTGT 2514
Qy 938 CTTCACTTTTGGAGGAG-----AAGCTCATTCGCGAATTTGGTAATTTGCTCTATCGA 991
Db 2515 ATTCCCTTTTACAGAGGGGTCTATCCAAAGGCCCTTCTTAGTGGGCCCTCTTTCTA 2574
Qy 992 CTCCTGGTAACTGGGAGAGAGAGTTTCCCATGATGGGCCCCACAGATGAACGTGTTATGT 1051
Db 2575 CCATCATCACTGGGAGCGGAGTTTGAATGTGGGCTCCAGACATGATGTGTAACCT 2634
Qy 1052 ATTTTGGCTGCGCAAGCTCGAGCAGTGTATCAGAGAACATGAGTTTACTTATCGAAG 1111
Db 2635 ATGTGGGTGACAAAGGACAGCCGTGCCATCATCCGAGAGAATGAGTTCTCTCTTTGAACA 2694
Qy 1112 ATCAAAAAGATCAAGAAAAGAAATCTGGACAAATAAGTACGAAAGCAAGCAAAA 1171
Db 2695 ATGCCATTCGTGGTGGCAAGAGGCTC-----CCGCATGAAGAAAGAGCAT 2742
Qy 1172 GAATCAAGTTTGATGTCTCTCCATCGTATGAGATGATCAACCTAGATTTCAGCAGTTC 1231
Db 2743 CTGTGAATTCATGTGCTGCTGACATCTATGAATTTGATCACCATTGACATGCTATTT 2802
Qy 1232 TAAACCAATTAAGTCGGAGTGATGATTGTTGATGAAGTCAAGTCAAGTGAAGAAATAGG 1291
Db 2803 TGGGCTCTATTGATTGGGCTGCTCATCGTGGATGAAGCCCATCGCTGAGAGAAATC 2862
Qy 1292 ATTCAAGCTGTTCTCTTCAATTGACACAGATATTCAGATTAACACCGCTATCTCTCAG 1351
Db 2863 AGTCTAAGTTCTTCCGGGTATTGAATGGTTACTCACTCCAGCACAAGCTGTTGCTGACTG 2922
```

Qy	1352	GAACACCACTT	CAGAACCAACTTGGATGAACTTTTCATGCTCATGCATTTTCTTGAATGCGG	1411
Db	2923	GGACACCATTTACAAAA	CAATCTGGAGAGTGTTCATCTGCTCAACTTCTTCTCACCCCGC	2982
Qy	1412	GGAAGTTTGAAGTTTGGAGAGTTCAGAGGAGGTTCAAGATATTAATCAAGAGGAGC	1471	
Db	2983	AGAGGTTTCCAATTTGGAAGGTTTTTGGAGGAGTTTGTGACATTTGCCAAGAGGAGACC	3042	
Qy	1472	AGATCTCAAGGTTGCACAAAATGTTGCTCCACATTTTGTCTCAGAAGGGTAAAAAAGAGC	1531	
Db	3043	AGATAAAAAAACTGCGATGACATGCTGGGGCGGCACATGTTGGCGGCTCAAAAGCCGATG	3102	
Qy	1532	TAATGAAGAAGCATGCCCCCCCCAAAAGAGAGCTCAATTTTTCGTGTGTGATCTCAGCAGTCTGC	1591	
Db	3103	TGTTCAAGAACATGCCCCCTCCAAGAGACGAACATAATTTGTGCGTGTGGAGCTGAGCCCTATGC	3162	
Qy	1592	AGAAAGAATATTAACAAGCTATTTTTTACCCTGTAATTAACAAGTATTGA	1648	
Db	3163	AGAAAGAAATACTACAAGTACATCTCACTCGAAATTTTGAAGCACTCAATATGCCGAGGTG	3222	
Qy	1649	GAGGTGCTCAAAATTTCCCTTAAATAACATTAATGATGGAATTAACGAAGAAGTATGTCGCATC	1708	
Db	3223	GTGGCAACCAAGGTGCTCTGCTGAATGTGTGATGGATCTTAAGAAGTGTCTGCAACCATC	3282	
Qy	1709	CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTTCACGACGCAAAATGAAGCTT	1759	
Db	3283	CATACCTTCTCCCTGTGGCTGCAATGAAGCTCCTAAGATGCCATATGGCATGTATGATG	3342	
Qy	1760	TCAACAACATTTTGGAGTCTTGTGGAAAGCTGCAACTTCTAGATAAATAATGATGGTCAAAAC	1819	
Db	3343	GCAGTGCCCTAATCAGAGACATCTGGGAAATATTGTCTGCTGCAGAAAAATGCTCAAGAACC	3402	
Qy	1820	TGAAGAGCAAGNACACAGAGTCTCTAATATACACACAGTTTCAGCATATCTCTGGACTTAC	1879	
Db	3403	TTAAGSAGGTTGGGCATCGTGACTCATCTTTTCCCAGATGACCAAGATGCTAGACCTGC	3462	
Qy	1880	TTGAAGACTACTGTACCATAAGAAATGGCAGTACGAGCGAAATTGATGGAAAGGTTGGCG	1939	
Db	3463	TAGAGGATTTCTTGGACATGAAGGTTATAATACGAACGCATCGATGGTGGAACTCACTG	3522	
Qy	1940	GAGCTGAGCGCAAAATACGATAGATCGGTTCAATGCCCCAAAATTTCAACAAGTTTGT	1999	
Db	3523	GGAACATGCGCAAGAGGCCATTGACCGCTTCAATGCACCCGGTGTCTCAGCAGTTCTGCT	3582	
Qy	2000	TTTTTGCTCTCCACAAGAGCTGTGGCTTAGGAATAAATCTTGCNAACGGCTGATACAGTAA	2059	
Db	3583	TCTTGTCTTTCCACTCGAGCTGGGGGCCCTTGGAAATCAATCTGGCCACTGCTGCACACAGTTA	3642	
Qy	2060	TCATTTATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC	2119	
Db	3643	TTATCTATGACTGATCTGNAACCCCATTAATGACATTCAGGCCCTTTAGCAGAGCTCACC	3702	
Qy	2120	GACTTGCCCAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCCACCATTTGAAG	2179	
Db	3703	GGATTGGCAAAATAAAGGTAAATGATCTACCGTTTTGTGACCCGTGCGTCACTGGAGG	3762	
Qy	2180	AAAGGATGATGCAATTGACTTAAAGAAAAATGGTTCTTAGAGCATCTTTGTTGTTGG-----	2234	
Db	3763	AGCGCATCAACGAGGTGGCAAGAAAGAAATGATGCTGAGCGCATCTAGTGGTGGCGCGCTG	3822	
Qy	2235	----GAAACTCAAAACACAAACATTAAATCAGGNAAGGTTAGATGACATCATCAGGTATG	2290	
Db	3823	GGCTGGGCTCCAGACTGGATCTATGTCCAAACAGGAGCTTTGATGATCTCTCAAAATTTG	3882	
Qy	2291	GATCAAAAGGAGCTTTTTTGTAGTCAAGATGATGAAGCAGGAAAGTCTGGAATAAATTCATT	2350	
Db	3883	GCATGAGGAACATAATTCAAGATGAAGCCACTGATGAGGAGGGAGACAAACAAGAGGGAG	3942	
Qy	2351	ATGATGATGGGCTTATAGACAAATTTGTTGA-----TCGTGATCTCTGGAGGCGAGAGG	2404	
Db	3943	AAGATAGCAGTGTATTCCACTACGATGATAAGGCCATTGAACGGCTCTCTAGACCGTAAACC	4002	

Qy	2405	AAGTCTCAGTGGATGATGAAGACAGAGAA	TGGATTCTTAAAGGCTTTCAAGGTCGGCTAATT	2465
Db	4003	AGGATCAGACTGAAGACACACAAATTCGAGGSCATGAATGAATATTTTGAAGCTCAATTCAAAG	4062	
Qy	2465	TTGCAATATATAGATGAAATGAGCGCAGCAGCAATTAGAGGCACACAGAGAGTGCCTGCTGAAA	2524	
Db	4063	TGGCCCAAGTATGTGTTACGGGAAGAAAGAAATGGGGAGAGNAGAGAGTTAGAACGGGAAA	4122	
Qy	2525	GCAAACTCTCAGCAGGCAATCTCTGATAGACAAAGTTATTGGGAAGAGTTGTTTAAAAAGATA	2584	
Db	4123	TCATTAACACAGGAAGAAAGTGTGATCCTGA---CTACTGGGAGAAATTGCTCGGCAC	4179	
Qy	2585	AATTGAGCTGCACACAGGCTGAGGAGCTTAACTCTCTTGTGAAAAAGAAAGAGAGTCGCA	2644	
Db	4180	ATTATGACGACGACGAAGAAGATCTAGCCCGAAATCTGGCGAAAGGAAAAGAAATCCGTA	4239	
Qy	2645	AGCAGTTGGTATCCATTGAAG---AAGATGATCTTCTGCTGGTTTGGAGATGTGAGCTCTG	2701	
Db	4240	AACAGGTCAACTACAATGATGGCTCCGACGAGACCGAGATTGGCAGGACGACACAGTCCG	4299	
Qy	2702	ATGAGAGTAAAGTTTATGAAGCTGAGTCAACAGATGCTGAAGCAGCAGGACAAAGGAGTTC	2761	
Db	4300	ACAAACAGTCCGATTAATCTAGTGGCTTCAGAGAAAGTGATGAAGACTTTGATGAACGTT	4359	
Qy	2762	AGACGGGTGCACGGCCGTACAGAAAGAAAGGGTCCGATATAATTTGSAACCAA-----	2812	
Db	4360	CAGAAAGCTCCCGTAGGCCAGTCGTAAGGGCTGCGGAATGATAAAGATAAGCCATTGC	4419	
Qy	2813	CTCGTTGATGGAAGGTGAGGGAGATCTTTACAGACTACTGGGTTTCAACCAGAGTCMAA	2872	
Db	4420	CTCCTCTGTTGGCCCGTGTGGTGGGAATATTGAAGTACTTGGTTTAAATGTCGTCGACG	4479	
Qy	2873	GGGCCATTTTGTACAGACTTTTCATGAGGTATCG-----AGCTGGCAATTTTGATT	2923	
Db	4480	GAAGAAGCTTCTTAATGCAATTAATGCGATATGATGCCACTCAGGATGCTTTTACTA	4539	
Qy	2924	GGAGGAGTTTGTTCCTCGCTTAAAGCAGAAAGCTTTGAAGAAATAAATGAATATGGAA	2983	
Db	4540	CCCAGTGGCTTGTAAAGAGACTCGCAGGCAATCAGAGAAAGAGTTCAAGGCATATGTCT	4599	
Qy	2984	TACTCTTCTTGAAGCACAATTGCTGAAGAAATAGACAGAAATTCCTCCAACCTTTTCAGATG	3043	
Db	4600	CTCTTTTTCATCGCGCATTTATGTGAGCCGGGCGAGATGGGGCTGAGACCTTTTGCTGATG	4659	
Qy	3044	GTGTGCCAAGGAAGGACTTTAGAATAGAAGATGCTTCTAGTCAGAAATTGCTCTCTGTATC	3103	
Db	4660	GTGTCCCCCGAAGGCGCTGTCTGCCGACATGTCTTACTAGAAATTTGGTTAATGTCCT	4719	
Qy	3104	TAGTTCAGGAGAGGTTGAAATTTGTAGAAGATCATCCAGG	3144	
Db	4720	TGATTCGACAGAGGTTCCAGGAGTTTGAACATGTTAATGGG	4760	
RESULT 11				
ADB48415				
ID	ADB48415 standard; cdna; 6475 BP.			
XX				
AC	ADB48415;			
XX				
DT	04-DEC-2003 (first entry)			
XX				
DE	Novel human cdna SEQ ID NO 325.			
XX				
KW	ss; cancer; neurodegenerative disease; human.			
XX				
OS	Homo sapiens.			
XX				
PN	US2003104529-A1.			
XX				
XX	05-JUN-2003.			
PD				
XX				
PF	04-JAN-2002; 2002US-00037270.			
XX				

PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 19-JUL-2000; 2000US-00620312.
XX
PA (ZHOU/) ZHOU P.
PA (TANG/) TANG Y T.
PA (LIUC/) LIU C.
PA (ASUN/) ASUNDI V.
PA (DRMA/) DRMANAC R T.
XX
XX Zhou P, Tang YT, Liu C, Asundi V, Drmanac RT;
XX WPI; 2003-678194/64.
XX
XX New polynucleotide, useful for treating diseases e.g., cancer or
XX neurodegenerative diseases.
XX
XX Claim 1; SEQ ID NO 325; 99pp; English.
XX
XX The invention relates to a polynucleotide comprising a sequence given in
XX the specification, or its mature protein-coding portion, or its
XX complement. The polynucleotide is useful for treating diseases e.g.,
XX cancer or neurodegenerative diseases and many others listed in the
XX specification. The present sequence represents a novel human cDNA. Note:
XX The sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format directly from USPTO
XX at seqdata.uspto.gov/sequence.html?docID=20030104529.
XX
SQ Sequence 6475 BP; 1760 A; 1528 C; 1822 G; 1365 T; 0 U; 0 Other;

Query Match 10.1%; Score 421.4; DB 8; Length 6475;
Best Local Similarity 51.9%; Pred. No. 1.4e-101;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;

QY 818 TACATCCATACAGCTTGAGGAGCTTAATTTTGGCGTCTCTCGTGTCAAAAACAGACGC 877
DB 2395 TGCACCCCTATCAATAGAGGGCTGAATGGTTCGCTCTCTCGGCTCAGGGCACTG 2454

QY 878 ATGTAATCTTGTGTGAATAGGACTAGCAAGACAATTCAGAACATTCGCCCTTTTAG 937
DB 2455 ACACCATCTTGGCTGATGAGATGGGCTTGGAAAACTGTACAGACAGCAGTCTTCCTGT 2514

QY 938 CTTTCACATTTTGGAGG-----AACCTCATTCGCGAATTTGGTAATTTGCTCTCTATCGA 991
DB 2515 ATTCCCTTTAAGAGGAGGGTCAITCCAAAGGCCCTTCTTAGTGAGCGGCCCTCTTTCTA 2574

QY 992 CTCTGCGTAATCTGGGAGAGAGTTTCCACATGGGCCCCACACATCAACGTTGTTATGT 1051
DB 2575 CCATCATCACTGGGAGGGAGTTTGAATGTGGGCTCCAGACATGATGTGTAACT 2634

QY 1052 ATTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTTACTTATCGAAAG 1111
DB 2635 ATGTGGGTGCAAGGACAGCGCTGCCATCATCCGAGAGATGAGTTCTCTTTGAAGACA 2694

QY 1112 ATCAAAAAGATCAAGAAAAGAAATCTGGCAAAATAGTAGGAAAGCAAGCAAAAAA 1171
DB 2695 ATGCCATTCGTGGTGGCAAGAGCCCTC-----CCGCATGAAGAAAAGAGGCAT 2742

QY 1172 GAATCAAGTTTGTCTCTCTCATCTGATGATGATGATCAACTAGATTTCAGCAGTTC 1231
DB 2743 CTGTGAATTCATCTGTCTGATCATCTTAATGATGATGATGATGATGATGATGATGAT 2802

QY 1232 TAAACCAATTAAGTGGGAGTGCATGATTTGTTGATGAAGGTTCATCGACTGAAAATAAGG 1291
DB 2803 TGGGCTCTATTGATTGGGCTGCCTCATCTGTTGATGAAGCCCATCGGCTGAAGAACATC 2862

QY 1292 ATTCAAAGCTTCTCTTTCATTCACACAGTATTCAGTAACCAACCGTATTTCTTCTGACAG 1351
DB 2863 AGTCTAAGTTCTTCGGGTATTGAATGGTTACTCACTCCAGCACAAAGCTGTGCTGACTG 2922

QY 1352 GAACACCACTTCAGAACAACTTGTGATGAACATTTTCATGCTCATGCAATTTCTTGATCGG 1411
DB 2923 GGACACCAATTAACAAAACAACTCGAAGAGTTGTTTCATCTGCTCAACTTTCTACCCCCG 2982

QY 1412 GGAAGTTTGAAGTTTGGAGGAGTTCCAGAGGAGTTCAAGATATTATTAATCAAGAGAGC 1471
DB 2983 AGAGGTTTCCAAATTTGGAAGGTTTGGAGGAGTTTGGAGGAGTTGCTGACATTTGCCAAGAGGAGC 3042

QY 1472 AGATCTCAAGGTTTGCAAAAATGTTGGCTCCCAATTTGCTCAGAAGGTTAAAAAAGACG 1531
DB 3043 AGATAAAAAAATGTCATGACATGCTGGGCGGCACATGTTGGCGCGCTCAAGAGCCGATG 3102

QY 1532 TAATGAAGACATGCCCCCAAAAAAGAGCTCAATTTTGGGTGTGATCTGAGCAGTCTGC 1591
DB 3103 TGTTCAGGAACATGCCCTCCAGACAGAACTAATTTGCGGTGTGGAGCTGAGCCCTATGC 3162

QY 1592 AGAAGCAATATTACAAAGCTATTTTACCCGTAATTTCAAGTATTGA---CAAAAAAGG 1648
DB 3163 AGAAGAAATACTACAAAGTACCTCCTCACTCGAAATTTTGAAGCACCTCAATGCCCGAGGTG 3222

QY 1649 GAGGTCTCAAAATTTCCCTTAATAACATTTATGATGAATTTACGAAAAGTATGCTGCCATC 1708
DB 3223 GTGGCAACAGGTGCTCTGCTGATGTTGGTGTGATCTTAAGAAGTGTGCAACCATC 3282

QY 1709 CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTCACGACGCAAAATGAAGCTT 1759
DB 3283 CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCTAAGATGCCCTAATGGCATGTATGATG 3342

QY 1760 TCAACAACATTTTGGAGTCTTGTGGAAGAGCTGCACTTCTAGATAAATGATGGTCAAC 1819
DB 3343 GCAGTCCCTTAATCAGAGCATCTGGGAAATTTATGCTGCTGCAGAAAATGCTCAAGAACC 3402

QY 1820 TGAAGACGCAAGGACACAGAGTCTTAATATACACAGTTTTCAGCATATGCTCGACTTAC 1879
DB 3403 TTAAAGAGGTGGGCATGCTGTAATCTTCTTCCAGATGACCAAGATGCTAGACTGC 3462

QY 1880 TTGAAGACTACTGTACCCCATAGAAATGGCAGTACGAGCGAATTTGATGGAAGGTTGGCG 1939
DB 3463 TAGAGGATTTCTTGAACATGAAGTTATAAATACGAACGATCGATGGTGAATCACTG 3522

QY 1940 GAGCTGAGCGGCAAAATACGATAGATCGGTTCAATGCCAAAAATTTCTAACAGTTTGT 1999
DB 3523 GGAACATCGGGCAAGAGGCCATTTGACCGCTTCAATGACCGGCTGCTCAGCAGTTCTGCT 3582

QY 2000 TTTTGTCTCCCAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA 2059
DB 3583 TCTTGTCTCCATCGAGCTGGGGGCTTGGAAATCAATCTGGCCACTGCTGACACAGTTA 3642

QY 2060 TCATTTATGACAGTGAATCCATCTGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
DB 3643 TTATCTATGACTGACTGGAACCCCAATAATGACATTCAGGCTTTAGCAGAGCTCACC 3702

QY 2120 GACTTGGCCAAACAAATAAGGTGATTTATAGGCTCATAAACCGAGGACCATTTGAAG 2179
DB 3703 GGAATGGCAAAAATAAAGAGTAAATGATCTACCGGTTTGTGACCGGCTGCTCAGTGGAGG 3762

QY 2180 AAAGGATGATGCAATTTGACTTAAAGAAATGTTCTAGAGCATCTTGTGTTGG----- 2234
DB 3763 AGCGCATCAGCAGGTGGCAAGAAAGAAATGATGCTGACGCAATGATGTTGGTGGCGCTG 3822

QY 2235 ----GAAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290
DB 3823 GGCTGGCTCCAGACTGGATCTATGTCCTCAAAACAGGAGCTTGTATGATATCTCTCAATTTG 3882

QY 2291 GATCAAGAGAGCTTTTGTGTAGTGAAGATGATGAACAGGAAAGTCTGGAATAATTCATT 2350
DB 3883 GCACCTGAGGAACCTATTCAAGGATGAAGCCACTGATGGAGGAGGAGACAAAGAGAGGGAG 3942

QY 2351 ATGATGATGGGCTATAGCAAAATTCCTGA-----TCGTGATCTCGTGGAGGAGAGG 2404
DB 3943 AAGATAGCAGTGTATTCCCACTACGATGAAGGCCATTTGAACGGCTGCTAGACCGTAACC 4002

QY 2405 AAGTCTCAGTGGATGATGAAGAGGAGAAATGGAATTTCTTAAAGGCTTTCAAGGGTGGCTAATT 2464
DB 4003 AGGATGAGACTGAAGACACAGAAATTCAGGGCATGATGATATTTGAGCTCATTCAAAG 4062


```
Db 2937 GCTGACAGGACCCCAATTGCAGAAATATCTGGAGGAGCTCTTCCATCTCCTGAACCTTCCT 2996
Qy 1404 TGATCGCGGGAAGTTTGGAGAGTTTGGAGAGATTCAGAGAGAGTTTCAAGATATTAATCA 1463
Db 2997 CACCCAGAGAGATTTAAACAATTTGGAGGCTTCTGGAGGAGTTTCTGACATATCCAA 3056
Qy 1464 AGAGGACGAGATCTCAAGGTTGACAAATGTTGGCTCCACATTTGCTCAGNAGGGTAAA 1523
Db 3057 AGAGGACGAGATCAAGAAATCTGCATGATTTGCTGGGGCCACATGCTCGGGAGACTCAA 3116
Qy 1524 AAAAGAGCTTAATGAAAGACATGCCCCCAAAAAGGAGCTCAATTTTGGCTGTGATCTGAG 1583
Db 3117 GGCAGATGCTTTAAGAAACATGCCAGCCAGACAGAGACTCATCGTTCCGGGTGGAGCTAAG 3176
Qy 1584 CAGTCTCAGAAAGAAATATTAAGAGCTATTTTACCCGTAATTTATCAAGTATTTGA---C 1640
Db 3177 CCCCATGCAGAAAGAAATACATAAATACATCTGACTCGAAATTTTGGAGCCTTGAATTC 3236
Qy 1641 AAAAAGGGAGGTGCTCAAAATTTCCCTTAATACATTTATGATGGAATTTACGAAAGTATG 1700
Db 3237 ACAGAGTGGTGGGAACACAGGTGTGCTGCTTAATATCATGATGGATCTTAAGAAAGTGTG 3296
Qy 1701 CTGCACTCTTATATGCTAGAGGCTGTG-----AGCCAGTTATTACAGACGCAAA 1751
Db 3297 CAACCATCAATCTTTTCCCGTGGCTGTATGGAGTCCCCAACTCCCGAGTGGGC 3356
Qy 1752 TGAAGCTTTTCAACAACCTTTTGGAGTCTTGTGGAAAGCTGCAATCTTAGATATAAATGAT 1811
Db 3357 TTATGAGGGTGGGCACTTTAATAGTCTGTGGGAAGCTCATGCTGCTCCAGAAGATGCT 3416
Qy 1812 GGTCAAACTGAAGAGCAAGGACACAGAGTCTTAATATACACAGTTTCAGCATATGCT 1871
Db 3417 GCGAAAGCTGAAGAGCAAGGACACCGAGTGTCTCATCTTCTCGCAGATGACCAAAATGTT 3476
Qy 1872 GGACTTACTTGAAGACTACTGTACCCATAAGAAATGCGAGTACGAGCGAATTCATGGAAA 1931
Db 3477 AGACTTCTTGAGACTCTTGTAGACTATGAGGCTACAGATATGAGGCGCATCGATGGTG 3536
Qy 1932 GGTGGCGGAGCTGAGCGGCAAAATACGATAGATCGGTTCAATGCCCAAAAATTTCTAACAA 1991
Db 3537 TATCACGGGTGCCCTGAGGAGGAGGCCATCGATCGGTTTAATGCTCCTGGGGCCCAACA 3596
Qy 1992 GTTTTGTGTTTGTCTTCCACAGAGCTGTGGCTTAGGATTAATCTTCGACCGGTGA 2051
Db 3597 ATTCTGCTTCTCTGTCTCCACCGAGCTGGGGGCTGGGCATCAATCTGGCCCATGCTGA 3656
Qy 2052 TACAGTAATCATTTATCACAGTGAATCGAATCTCATGCTGATCTTCAAGCAATGGCTAG 2111
Db 3657 CACTGTCAATCATCTTTGATCTGACTGGAACCCCAATAAGCAATCCAGGCTTTAGCCG 3716
Qy 2112 AGCTCATCGACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCAC 2171
Db 3717 GGCTCATCGGATTTGGCCAGGCCAACAAAGTGATGATTTACCGGTTTGTGACTCGCGGTC 3776
Qy 2172 CATTGAAGAAGGATGATGCAATTTGACTAAAGAAATAGTTTCTAGAGCATCTTGTGTT 2231
Db 3777 AGTGAAGAGCGAATCACAAAGTGGCCAAAGAGAAAGATGATGCTGACACACCTTGTGTT 3836
Qy 2232 TGG-----GAAACTCAAAACACAAACATTAATCAGMAGAGTTAGATGACATCAT 2282
Db 3837 GCGGCCCTGGGCTGGGCTCCAGGCGAGGCTCATGCTCCAAGCAGGAGCTTCAGCAGATCT 3896
Qy 2283 CAGGTATGGATCAAGAGAGCTTTTGTAGTGAAGATGATGAAGCAGGAAAGTCTG---- 2338
Db 3897 CAATTTGGCACTGAAGAGCTATTCAAGGATGAAACGAGGGGAGAAACAGGAGGAGGA 3956
Qy 2339 -----GAAAAATTCATTATGATGATGCGGTATAGACAAATTTGCTTGTATCTCGT 2393
Db 3957 CAGCAGTGTGATTCATTATGACAAATGAGGCGCATCGCTCGGCTGTTGGACCGGAACACGGA 4016
Qy 2394 GGAGGACAGGAAGTCTCAGTGGATGATGAAGAGGAGATGGATCTTTAAGGCTTTCAA 2453
Db 4017 TGCAACTGAGGA-----CACTGACGTGCAGAACATGAATGAGTATCTCAGTCTCTTCAA 4070
```

```
Qy 2454 GGTGGCTAAATTTTGAATATATATAGATCAAAATGAGGCAGCAGCATTTAGAGCCACAGAGAT 2513
Db 4071 GGTGGCACAGTACGTCGTCGGGAAGAGACAAGATTTGAGGAATTTGAGCGAGAGATCAT 4130
Qy 2514 CGCTGCTGAAAAGCAAAATCTTTCAGCAGGCAATTCGTATAGAGCAAGTTATTGGGAAGAGTT 2573
Db 4131 CAAGCAGGAGGAGATGTGGACCTGAC-----TACTGGGAGAGCT 4172
Qy 2574 GTTAAAGATAAAATTTTGTAGCTGCACAGCTGAGGAGCTTAAATGCTTCTTGGAAAAAGGAA 2633
Db 4173 GCTGAGGCTACTATATGAGCAACAGCAGGAAGACCTAGCCGGAATCTAGGCAAGGCAAA 4232
Qy 2634 GAGAACTCCGAAGCAGTTGGTATCTCATTGAAGAGATGATCTTGTGCTTGTGGTATGAAATGT 2693
Db 4233 GCGGGTTCGCAAGCAAGTTTAACATACATGATGCTCTAGGAAGAGCAAGAACACCAAGTC 4292
Qy 2694 GAGCTCTGATGAGGATGAAAGTTTATCAAGCTGAGTCAACAGATGGTGAAGCAGCAGACA 2753
Db 4293 AGAGTACTCGTGGGTTTCAAGGAGGAGATGAAGACTTCGATGAACGCTCTGAAGGGCG 4352
Qy 2754 AGGAGTTTCAGACGGGTCCGACGGCCGTACAGAAAGAAAGGGTCCGCTAATTTTGAACCAAC 2813
Db 4353 TAGACAGTCAAGAGGAGCAGCTCCGGAA-----TGAGAAAGATAAGCCACTGCC 4400
Qy 2814 TCCGTTGATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACAGAGATCAAG 2873
Db 4401 TCCACTGCTGGCCCGAGTCCGGGGCAACATTGAGGTGCTGGGCTTCAACACCCGCTCAGCG 4460
Qy 2874 GGCCATTTTGTACAGACTTTTGTATGAGGTATGG-----AGCTGGCAATTTTGTATTG 2924
Db 4461 GAAGGCTTTCCTCAATGCTGTGATGCTGGGGGATGCCACACAGATGCTTTCACCAC 4520
Qy 2925 GAAGGAGTTTGTTCCTCGCTTAAAGCAGAAAGACCTTTTGAAGAAATAAATGAATATGGAAT 2984
Db 4521 ACAGTGCCTGTCGGGACCTGAGGGGCAAGACTGAGAGGAGTTTAAAGSCTATGTGTC 4580
Qy 2985 ACTTCTTTGAAGCAATTTGCTGAAGAAATAGACAGAAATTTCCAAACCTTTTCAGATGG 3044
Db 4581 TTTGTTTCATCGCCCATCTGTGTGAGCCTGGGGCAGACGGCTCTGAACCTTTTCCCGATGG 4640
Qy 3045 TGTGCCCAAGGAGGACTTAGAATAGAATGTTCTAGTCAGAAATTTGCTTCTCTGATACT 3104
Db 4641 GGTCCCTCGGAGGAGTCTGATGCGCCAGAGGTTGTACCCGCAATTTGGAGTCAATGCTCT 4700
Qy 3105 AGTTCAGGAGAGGTGAA 3122
Db 4701 CGTCAAAAGAGGTGCA 4718
```

RESULT 13

AAD54629

ID AAD54629 standard; DNA; 6331 BP.

XX AAD54629;

XX AAD54629;

XX AAD54629;

DT 26-JUN-2003 (first entry)

XX Human chromodomain helicase DNA binding protein (CHD) encoding DNA #4.

XX Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;

XX chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;

XX cell proliferative disorder; chromatin organisation modifier domain;

XX cytosolic; ds.

XX Homo sapiens.

XX WO200298899-A2.

XX 12-DEC-2002.

XX 03-JUN-2002; 2002WO-US017466.

XX

PR 05-JUN-2001; 2001US-0296076P.
PR 10-OCT-2001; 2001US-0328605P.
PR 22-OCT-2001; 2001US-0338733P.
PR 15-FEB-2002; 2002US-0357253P.
PR 15-FEB-2002; 2002US-0357600P.
XX (EXEL-) EXELIXIS INC.
XX
XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
XX Lioubin MN;
XX WPI; 2003-156840/15.
XX
XX Identifying a candidate p53 pathway-modulating agent as therapeutic
XX targets for disorders related to defective p53 function e.g. cancer by
XX contacting an assay system having purified CHD polypeptide or nucleic
XX acid, with a test agent.
XX
XX Disclosure; Page 51-54; 278pp; English.
XX
XX The present invention relates to a method for identifying candidate p53
XX pathway modulating agents. The method involves contacting an assay system
XX comprising purified chromatin organisation modifier (chromo) domain
XX helicase DNA binding proteins (CHD), nucleic acids, their functionally
XX active fragments or derivatives, with a test agent under conditions
XX where, but for the presence of the test agent, the system provides a
XX reference activity. The methods are useful for identifying modulators of
XX the p53 pathway as therapeutic targets for disorders associated with
XX defective p53 function, such as angiogenic disorders, apoptotic disorders
XX or cell proliferative disorders, e.g. cancer. The modulators are useful
XX as research reagents, diagnostics and therapeutics. The invention is also
XX useful in gene therapy. The present sequence is human CHD DNA
XX
XX Sequence 6331 BP; 1720 A; 1552 C; 1819 G; 1240 T; 0 U; 0 Other;
XX
XX Query Match 10.1%; Score 420.8; DB 7; Length 6331;
XX Best Local Similarity 52.3%; Pred. No. 2e-101;
XX Matches 1233; Conservative 0; Mismatches 1032; Indels 93; Gaps 10;
XX
XX 810 AGGCTGTGATCATCATCAGCTTGGAGGACTTAATTTTGGCGTTCTCGTGGTCAAA 869
XX
XX 2409 AGGCACCTCGACATGATATCATGTTGGAAGGCTGAACTGGCTACGCTTCTCTGGGCCA 2468
XX
XX 870 ACAGAGCATGTAATCTTCTGATGAATGGGACTAGGCAAGACAATTCAAACATTCG 929
XX
XX 2469 GGGCACTGACACCATTTAGCTGATGATGAGTGGGCTAGGCAAGACCATACAAACATCGT 2528
XX
XX 930 CTTTATAGCTTCACTTTTGGAGAGAACCTCA-----TTCCGCAATTTGGTAAATGCTCC 983
XX
XX 2529 CTTCCTCTACTCACTACAAAGGAGGGGCCACAAAAGGTCCTTCTGTTGAGTGCCCC 2588
XX
XX 984 TCTATCGACTCGGTAACTGGGAGAGAGATTTGGCCACATGGGGCCCAACAGATGAAGCT 1043
XX
XX 2589 ACTCTCTACCATCATTAATCTGGGAGCGGGAGTTCCAGATGTGGGACCCCAATTTCTATGT 2648
XX
XX 1044 GGTATGATTTTGGCACTGCGCAAGCTCGAGAGAGTTATCAGAGAAATGATGATGTTTACTT 1103
XX
XX 2649 GGTGACATACAGGGTGAAGGACAGCGGGGCCATCATCTCGTGAAGATGAATTTCTCTT 2708
XX
XX 1104 ATCGAAAGATCAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGATAGCGAAAGCAA 1163
XX
XX 2709 TGAGGACAAATGCCATCAAGGGGGCAAGAAAGCTTTTAAAGTAAAG----- 2756
XX
XX 1164 GCAAAAAAGAAATCAAGTTTGTATGTCCTCTCACATCGTATGATGATCAACCTAGATTC 1223
XX
XX 2757 GGAGGCAAGGTGAAGTTTCCATGTTTCTCTGACATCGTATGAGTGTATCACCATTGATCA 2816
XX
XX 1224 AGCAGTTCTAAACCAATTAAGTGGAGTGCATGATTTGATGAAAGTCAATCCACTGAA 1283
XX
XX 2817 GGCAGCACTTGGTTCATCCGCTGGGCTGTCCTGTTGGTAGATGAGGCCCATCCACTCAA 2876
XX
XX 1284 AAATAAGGATTCAAAGCTGTTCTCTTCAATTGACACAGTATTCAAAGTAACCAACCGTATTC 1343
XX
XX

Db 2877 GAACAACCAAGTCCCAAGTTTTTTCAGGGTCTCTCAATGGTTTACAGATAGATCATAAAGTTGCT 2936
Qy 1344 TCTGACAGGAGAACCACTTTCAGAACCACTTGGATGAACCTTTTCATGCTCATGATTTCT 1403
Db 2937 GCTGACAGGAAACCCCATTTGAGAAATATCTGGAGGAGCTCTTCCATCTCTGAACTTCT 2996
Qy 1404 TGATGGGGGGAAGTTTGGAAAGTTTGGAGGAGTTTCCAGGGAGGAGTTCAAGATATTAATCA 1463
Db 2997 CACCCAGAGAGATTTAAACRACTTGGAGGCTTCTCTGGAGGAGTTTCTGACATATCAA 3056
Qy 1464 AGAGGAGCAGATCTCAAGGTTGCAAAAATGTTGGTCTCCACATTTGCTCAGAAAGGTAAA 1523
Db 3057 AGAGGACCAAGATCAAGAAACTGCATATTTGCTGGGGCCACACATGTCGGAGACTCAA 3116
Qy 1524 AAAAGAGCTAATCAAGAGACATGCCCCCAAAAGAGGAGCTCATTTTGGTCTGATCTGAG 1583
Db 3117 GGCAGATGCTTTTAAAGAACATGCCCAAGACAGAGCTCATCTGCTGGGTGGAGCTAAG 3176
Qy 1584 CAGTCTGCAGAAAGAAATATTACAAAGCTATTTTACCCGCTAATTTATCAAGTATTGA---C 1640
Db 3177 CCCCATGCAGAGAAATATCTACAAATACATCTCTGACTCGAAATTTTGAGGCTTTGAATTC 3236
Qy 1641 AAAAAAGGGAGGTGCTCAAAATTTCCCTTAATAACATATGATGGAATTAGCAAAAGTATG 1700
Db 3237 ACGAGGTGGTGGGAACCAAGGTGCTGCTTTAATATATCATGATGATGATCTTTAAGAAGTCTG 3296
Qy 1701 CTGCCATCTTATATGCTAGAGGCTGTTG-----AGCCAGTTATTTCACGACGCAAA 1751
Db 3297 CAACCATCCATACCTTTTCCCGTGGCTGCTATGGAGTCCCCCAAACTCCCAAGTGGGGC 3356
Qy 1752 TGAAGCTTTTCAAAACAACTTTTGGAGTCTTGTGGAAAGCTGCACACTTCTTAGATAAAATGAT 1811
Db 3357 TTATAGGGTGGGCACTTATTAAAGTCGCTGGGAAGCTCATGCTGCTCCAGAAGATGCT 3416
Qy 1812 GGTCAAACTCAAAAGACAGGACACAGAGTCTTAATAATATACACAGTTTTCAGCATATGCT 1871
Db 3417 GCGAAAGCTTGAAGAGCAAGGACACCGAGTGTCTCATCTTCTCCAGATGACCAAAATGTT 3476
Qy 1872 GGACTTACTTGAAGACTACTGTACCCATAGAAATGGCAGTACGAGCGAATTCATGAAA 1931
Db 3477 AGACTTGTCTTGGAGCTTCTTAGACTATGAAGCTCAAGATGAGGCGCATCATGTTGGTGG 3536
Qy 1932 GGTGGCGGAGCTGAGCGGCAAAATACGCTAGATCGGTTCATTCGCAAAAATTTCTAACAA 1991
Db 3537 TATCAGGGTGCCTGAGGCGAGGCGCATCGATCGTTTAAATGCTCTCTGGGCCCCAACA 3596
Qy 1992 GTTTTGTGTTTGTCTTCCACAGAGCTGTGTGCTTAGGAAATAAATCTTTCACAGCGTGA 2051
Db 3597 ATTCTGCTTCTCTGCTCCACCGAGCTGGGGGCTGGGCAATCAATCTGCGCACTGCTGA 3656
Qy 2052 TACAGTAATCATTTATGACAGTCACTGGAATCTCTCATGCTGATCTTCAAGCAATGGCTAG 2111
Db 3657 CACTGTATCATCTTTGATTTCTGACTTGGAAACCCCAATATGACATCAGGCGCTTTAGCCG 3716
Qy 2112 AGCTCATCTGATTTGGCCAAAACAAATAGGTGATGATTTATAGGCTCATAAACCGAGCAC 2171
Db 3717 GGCTCATCGATTTGGCCAGGCGCAAAAGTATGATTTTACCGGTTTGTGCTGCGCGTCT 3776
Qy 2172 CATTTGAAGAAAGGATGATGCAATTTGACTAAAAAGAAATGGTTCTTAGAGCATCTTGTGCT 2231
Db 3777 AGTGAAGAGCGAATCACACAGTGGCCAAAGAGAAAGATGATGCTGACACACCTGGTTGT 3836
Qy 2232 TGG-----GAAACTCAAAACACAAAACATTAATCAGAGAGAGTTAGATGACATCAT 2282
Db 3837 GCGGCTTGGGCTGGGCTCCAAAGCGAGGCTCCATGTCCAAAGCAGGAGCTTGACGACATCT 3896
Qy 2283 CAGTATGAGTCAAAAGAGCTTTTGTCTAGTGAAGATGATGAAGCAGGAGAAAGTGTG--- 2338
Db 3897 CAAATTTGGCACTCAAGAGCTATTCAAGGATGAAACGAGGGGAGAAACAAAGGAGGAGGA 3956
Qy 2339 -----GAAAAATTCATTTATGATGATGCGGCTATAGACAAATTCCTGATCGTATCTCGT 2393
Db 3957 CAGCAGTGTGATTTCAATTTATGACAAATGAGGCCATCGCTCGGCTGTTGGACCGGAACCAAGGA 4016

Qy	2394	GGAGGCAGAGGAAGTCTCAGTGGAGTATGAAGAGGAAGTGGATCTTTAAAGGCTTTCAA	2453
Db	4017	TGCAACTGAGGA-----CACTGACGCTGCAGAACATGAATGAGTATCTCAGCTCCTTCAA	4070
Qy	2454	GGTGGCTAAATTTTGAATATATAGATGAAATGAGCGACGACATTAGAGCCAAGAGAGT	2513
Db	4071	GGTGGCACAGTACGTCGTGCGGGAAGAAGACAAGATTGAGGAAATTCAGCGAGAGATCAT	4130
Qy	2514	CGCTGCTGAAAGCAAATCTTCAGCAGGCAATTCGTATAGACGAAGTTATTGGGAAGAGTT	2573
Db	4131	CAAGCAGGAGGAGATGTGACCTTGAC-----TACTGGGAGAGCT	4172
Qy	2574	GTTAAAAAGATAAATTTGAGCTGCAACAGGCTGAGGAGCTTTAATGCTCTTCGGAAAAAGGAA	2633
Db	4173	GCTGAGGCATCATCTATGAGCAACAGCAGGAGAACCTAGCCCGGAATCTAGGCAAGGSCAA	4232
Qy	2634	GAGAAATGCGCAAGCAGTTGGTATCCATTGAGAAAGATGATCTTGCTCGTGGTTTGGAGATGT	2693
Db	4233	GCGGGTTTCGCAAGCAAGTTAACTACAATGATGCTGCTCAGGAAGACCAAGACAACCAAGTC	4292
Qy	2694	GAGCTCTGATGGAGATGAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACA	2753
Db	4293	AGAGTACTCTCGTGGGTTTCAGAGGAGGAGGATGAAGACTTCGATGAACGCTCTGAAGGGCG	4352
Qy	2754	AGGAGTTTCAGACGGGTGCAGCGCCGTACAGAAAGAGGGTCGCGATAAATTTGGAACCAAC	2813
Db	4353	TAGACAGTCAAGAGCGAGCTCCGGAA-----TGAGAAAGATTAAGCCACTGCCC	4400
Qy	2814	TCGGTTGATGGAAGGTGAGGGGAGATCTTTTCAGAGTACTGGGTTTCAACAGAGTCAAAG	2873
Db	4401	TCCACTGCTGCCGAGTCGGGGCAACATTGAGGTCTGGGCTTCAACACCCGTCAAGC	4460
Qy	2874	GGCCATTTTGTACAGACTTTGATGAGGTATGG-----AGCTGGCAATTTTGATG	2924
Db	4461	GAAGGCTTCTCTCAATCTGTATGTCGTGGGGATGCCACACAGAGATGCTTCCACC	4520
Qy	2925	GAAGGAGTTTGTCTCGCTTAAAGCAGAGACCTTTTGAAGAAATAATGAATATGGAAT	2984
Db	4521	ACAGTGGCTGTGCGGACCTTGAGGGGCAAGACTGAGAAGGAGTTTAAGGCCCTATGTGTC	4580
Qy	2985	ACTCTCTTGAAGCACATTCGTGAAGAAATAGACGAGAAATTCCTCAACCTTTTCAGATGG	3044
Db	4581	TTTGTTTCATGGCCATCTGTGTGAGCCTGGGGCAGAGGCTCTCGAAACCTTTGCCGATGG	4640
Qy	3045	TGTGCCCAAGGAAGCACTTAGAATPAGAAGATGTTTCTAGTCAGAAATTCCTCTCTGATACT	3104
Db	4641	GGTCCCTCGGAGGGACTGAGTCGCCAGAGGTGTTGACCCGATTCGGATGCTGCTCT	4700
Qy	3105	AGTTCAGGAAAGGTGAA	3122
Db	4701	CGTCAAAAAGAGGTGCA	4718

RESULT 14
ACD13374

ACD13374
ID ACD13374 standard: cDNA: 6331 BP.

AC ACD13374:

PT 13-AUG-2003 (first entry)

Human DNA encoding a p53 modifier. SEQ ID 44.

Human; ss; gene; p53 modifier; cytostatic; cancer; cytostatic;
 antiangiogenic; antiapoptotic; p53 pathway; breast cancer; colon cancer;
 kidney cancer; lung cancer; ovarian cancer; angiogenesis; cell cycle;
 apoptotic disorder; cell proliferation disorder.

OS Homo sapiens.

XX
XX

PNXX

12-DEC-2002.

03-JUN-2002; 2002WO-US017382.

05-JUN-2001: 2001US-0296076P.

10-OCT-2001; 2001US-0328605P.

15-FEB-2002; 2002US-0357253P.

(EXEL-) EXELIXIS INC.

Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;

WPI; 2003-156859/15.

P-PSDB; ABO07199.

Identifying modulators of the p53 pathway for use in treating apoptotic or cell proliferation disorders, comprises screening for agents that modulate activity of a human ortholog of genes that modify the p53 pathway in *Drosophila*.

Example 2: page 269-271: 678pp: English.

The invention relates to identifying (M1) a candidate p53 pathway modulating agent, by contacting an assay system comprising a purified HM polypeptide (human orthologue of genes that modify the p53 pathway in *Drosophila*) or nucleic acid with a test agent under conditions, where but for the presence of the test agent, the system provides a reference activity, and detecting a test agent-biased activity of the assay system. Also included are modulating (M2) a p53 pathway of a cell (comprising contacting a cell defective in p53 function with a candidate modulator that specifically binds to a HM polypeptide comprising an HM amino acid sequence, where p53 function is restored), modulating (M3) a p53 pathway in a mammalian cell (comprising contacting the cell with an agent that specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4) a disease in a patient (comprising: (a) obtaining a biological sample from the patient; (b) contacting the sample with a probe for HM expression; (c) comparing the results with a control; and (d) determining whether the comparison indicates a likelihood disease). (M1) is useful for identifying modulators of the p53 pathway. A probe for HM expression is useful for diagnosing breast, colon, kidney, lung and ovarian cancer, in a patient, where the cancer has greater than 25 % expression level. Modulators identified by (M1) are useful in a variety of diagnostic and therapeutic applications, where disease or disorder prognosis is related to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell proliferation disorders (e.g. cancer). Another two new methods (M2 and M3) are useful for modulating the p53 pathway of a cell, thus restoring the p53 function of the cell, so that the cell undergoes normal proliferation or progression through the cell cycle. (M2) and (M3) are also useful for treating defects in the p53 pathway such as angiogenic, apoptotic or cell proliferation disorders. The present sequence is an HM nucleic acid encoding a p53 pathway modifying protein.

Sequence 6331 BP: 1720 A: 1552 C: 1819 G: 1240 T: 0 U: 0 Other:

Query Match 10.1%; Score 420.8; DB 7; Length 6331;
Best Local Similarity 52.3%; Pred. No. 2e-101;
Matches 1233; Conservative 0; Mismatches 1032; Indels 93;

Qy	810	AGCGTTGTTACATCCATACCAGCTTGAGGGAGCTTAATTTTTCGGTCTCGTGGTCAAA	869
Db	2409	AGGACCCCTGCACATGTTTCAGTTTGGAAAGGCTGACTGGCTTCGTTCTCTGGGCCCA	2466
Qy	870	ACAGAGCGCATGTAACTCTTCTGTAGTAAATGGGACTAGGGCAAGACAATTCAAAGCATTTGC	929
Db	2469	GGGCATCTGACACCACTTCTAGCTGATGAGATGGGCTAGGCAAGACCATACAAACCATCGT	2522
Qy	930	CTTTTTCAGCTTTCACCTTTTTCAGAGAGAACCTCA-----TTCGGCATTTGGTAAATTTGCTTC	983
Db	2529	CTTCTCTACTACTCTTACAAGAGAGGGCCACACAAAAGTTCCTTCTCGTGGTGGTGGCCCC	2588

Qy	1871	TGGACTTACTTGAAGACTTACTGTACCCATAGAAATGCGAGTACGAGCAATTTGATGAA	1930
Db	3759	TAGACTTGTCTTGAGCACTTCTTAGACTATAGAGCTACAGTATAGCGCATCGATGGTG	3818
Qy	1931	AGGTTGGCGGAGCTGAGCGCAATACGCTAGATCGGTTCAATGCCAAAAAATTTCTAACA	1990
Db	3819	GTATCAGGGTGCCTCGAGCGAGGCCATCGATCGGTTAATGCTCCTGGGCCCAAC	3878
Qy	1991	AGTTTGTGTTTGTCTCTCCACAGAGCTGTGTGCTTAGGATAAATCTTGCAACGGCTG	2050
Db	3879	AAITCTGCTCTCTCTGTCCACCGGAGCTGGGGCCCTGGGCATCAATCTGGCCACTGCTG	3938
Qy	2051	ATACAGTAATCATTTATGACAGTGAATCTCATGCTGATCTTCAAGCAATGCTA	2110
Db	3939	ACATGTCAATCATCTTTGATTTGATTTGATTTGGAACCCCAATATGACATCCAGGCTTTAGCC	3998
Qy	2111	GAGCTCATC-GACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGC	2169
Db	3999	GGGCTCATCGGATTTGGGCCAGGCCCAACAAAGTGATGATTTACCGGTTTGTGACTCGCGC	4058
Qy	2170	ACCATGGAAGAGATGATGCAATTTGACTTAAAGAAATGGTTCTTAGAGCATCTTGT	2229
Db	4059	TCAGTGGAGAGCGAATCACACAAGTGGCCCAAGAGAAGATGATGCTGACACACCTCGTT	4118
Qy	2230	GTTCG-----GAAACTCAAAACACAAACATTAAATCAGAGAGAGTTAGATGACATC	2280
Db	4119	GTGGCGCTCGGCTGGGCTCAAGCGAGGCTCAATGTCCAAGCAGGAGCTTGACGACATT	4178
Qy	2281	ATCAGGTATGGATCAAGGAGCTTTTGTCTAGTGAAGATGATGAAGCAGGAAAGTCTG--	2338
Db	4179	CTCAAAATTTGGCACTGAAGAGCTATTCAAGGATGAAACAGAGGGGGAGAACAGGAGGAG	4238
Qy	2339	-----GAAAAATTCATTATGATGATCGGCTATAGACAAATTTGCTTGATCGTGATCTC	2391
Db	4239	GACAGCAGTGTGATTCATTATGCAATGAGGCCATCGCTCGGCTGTTGGACCCGGAACCA	4298
Qy	2392	GTGGAGGCAGAGGAAGTCTCAGTGGATGATCAAGAGGAGATGATTTCTTAAAGGCTTTC	2451
Db	4299	GATGCACTGAGGA-----CACTGACGTGCAGAACATGATGATATCTCAGCTCCTTC	4352
Qy	2452	AAGTGGCTAATTTGTAATATATAGATGAAATGAGCGAGCAGCATTAGAGGCACAGAGA	2511
Db	4353	AAGGTGGCACAGTAGCTCGTGGGGAAGACAAAGATTGAGGAAATTGACGAGAGATC	4412
Qy	2512	GTGCTGCTGAAAGCAATCTTCAGCAGGCAATTTCTGATAGCAAGTATTGCGGAAGAG	2571
Db	4413	ATCAAGCAGGAGGAGAAATGTGGACCCCTGAC-----TACTGGGGAAG	4454
Qy	2572	TTGTTAAAGATAAATTTGAGCTGCACCAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGG	2631
Db	4455	CTGCTGAGGCATCACTATGACCAACAGCAGGAAGACCTAGCCCGGAATCTAGGCAAGGC	4514
Qy	2632	AAGAGAGTCCGAAAGCTTGGTATCCATTCAGAAAGATGATCTTGCTGGTTTGGAAAGAT	2691
Db	4515	AAGCGGGTTCGCAAGCAAGTTAACTACAATGATGCTGCTCAGGAAGACCAAGACAACAG	4574
Qy	2692	GTGAGCTCTGATGAGGATGAAGTTATGAAGCTGAGTCAACAGATGTTGAGCAGCAGGA	2751
Db	4575	TCAGAGTACTCGGTGGGTTTCAGAGGAGGAGGATGAAGACTTCGATGAACGCTCTGAAGGG	4634
Qy	2752	CAAGGAGTTACAGCGGTCGACGGCGTACAGAGAAAGGTCGCGATAATTTGGAACCA	2811
Db	4635	CGTAGACAGTCAAGAGGCGAGCTCGGAA-----TGAGAAAGATAAGCCACTG	4682
Qy	2812	ACTCGTTTGTAGGAGGTGAGGGAGATCTTTTACAGATATGCGGTTTCAACAGAGTCAA	2871
Db	4683	CCTCCACTGCTGGCCGAGTGGGGGCAACATTGAGGTGCTGGGCTTCAACACCCGTCAG	4742
Qy	2872	AGGGCCATTTTGTACAGACTTTGATGAGGTATGG-----AGCTGGCAATTTTGTAT	2922
Db	4743	CGGAAGGCTTTCTCTCAATGCTGTGATGCTGGGGGATGCCAACACAGGATGCTCTTACC	4802

Search completed: September 18, 2004, 20:33:46
Job time : 1494 secs

Qy	2923	TGGAAGGAGTTTGTCTCTCGCTTTAAAGCAGAGACCTTTGAAGAAATAAATGAATATGGA	2982
Db	4803	ACACAGTGGCTGTGCGGGACCTGAGGGCAAGACTGAGAGAGGAGTTTAAGGCTATGTG	4862
Qy	2983	ATACTCTTTTGAAGCACATTGCTGAAGAAATAGACGAGNAATTTCTCCAACCTTTTCAGAT	3042
Db	4863	TCCTTGTTCATGCGCCATCTGTGTGAGCCTTGGGGCAGACGGCTCTGAAACCTTTGCCGAT	4922
Qy	3043	GGTGTGCCCAAGGAGACTTTAGAAATAGAAAGATGTTCTAGTCAGAAATTTGCTCTTCTGATA	3102
Db	4923	GGGTCCCTCGGAGGAGCTGAGTCCGACAGGTTGTGACCCGCAATTTGGAGTCATGCT	4982
Qy	3103	CTAGTTTCAGGAGAAAGTGAA	3122
Db	4983	CTCGTCAAAAAGAGGTGCA	5002

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	421.4	10.1	6328	4	US-08-913-832A-1	Sequence 1, Appli
2	421.4	10.1	6328	4	US-09-249-181A-1	Sequence 1, Appli
3	421.4	10.1	6328	4	US-09-158-707-1	Sequence 1, Appli
4	421.4	10.1	6475	4	US-09-620-312D-325	Sequence 325, App
5	285.4	6.8	5837	4	US-09-976-594-75	Sequence 75, Appl
6	213.6	5.1	2512	4	US-09-702-705-1802	Sequence 1802, Ap
7	213.6	5.1	2512	4	US-09-736-457-1802	Sequence 1802, Ap
8	213.6	5.1	2512	4	US-09-671-325-1802	Sequence 1802, Ap
9	112.2	2.7	8916	4	US-09-579-181-11	Sequence 11, Appl
10	112.2	2.7	9354	4	US-09-579-181-10	Sequence 10, Appl
11	103.2	2.5	1868	4	US-09-023-655-429	Sequence 429, App
12	103.2	2.5	575	3	US-09-328-111-553	Sequence 553, App
13	100.2	2.4	5386	4	US-09-535-008-60	Sequence 60, Appl
14	100.2	2.4	5468	4	US-09-535-008-66	Sequence 66, Appl
15	100.2	2.4	5471	4	US-09-535-008-1	Sequence 1, Appli
16	100.2	2.4	5471	4	US-09-535-008-62	Sequence 62, Appl
17	100.2	2.4	5471	4	US-09-535-008-74	Sequence 74, Appl
18	100.2	2.4	5480	4	US-09-535-008-70	Sequence 70, Appl
19	100.2	2.4	5564	4	US-09-535-008-68	Sequence 68, Appl
20	100.2	2.4	5567	4	US-09-535-008-64	Sequence 64, Appl
21	100.2	2.4	5573	4	US-09-535-008-76	Sequence 76, Appl
22	100.2	2.4	5576	4	US-09-535-008-72	Sequence 72, Appl
23	90.6	2.2	1220	4	US-09-205-258-140	Sequence 140, App
24	89.2	2.1	3436	3	US-09-276-531-60	Sequence 60, Appl
25	88.2	2.1	345	4	US-09-016-434-15	Sequence 15, Appl
26	79.8	1.9	3593	4	US-09-404-627-3	Sequence 3, Appli
27	79.8	1.9	4205	4	US-09-404-627-1	Sequence 1, Appli

Db 2477 ATGCCATTCGTGGTGGAAGAGCCCTC-----CCGCATGAAGAAAGAGGCAT 2524
Qy 1172 GAATCAAGTTTGATGTCTCTCTCATCTGATGATGATGATCAACCTAGATTCAGCAGTTC 1231
Db 2525 CTGTGAATTCATGTGCTGCTGATCATCTCTATGAATTCATCACCATTGACATGGCTATTT 2584
Qy 1232 TAAACCAATTAAGTGGGAGTGATGTTGTTGATGAAGGTCATCGATCGAATAAAGG 1291
Db 2585 TGGGCTCTATTGATGGGCGCTGCTCATCTGCTGGATGAAGCCCATCGGCTGAAGAACATC 2644
Qy 1292 ATTCAAAGCTGTTCTCTTCATTGACACAGTATTCAAAGTAACACCGTATTTCTTGACAG 1351
Db 2645 AGTCTAAGTTCTTCGGGTATTGAATGTTTACTCACTCCAGCACAAAGCTGTTGCTGACTG 2704
Qy 1352 GAACACCACTTCAGAACAACTTTGATGAACTTTTCATGCTCATGCTATTTCTTGATCGGG 1411
Db 2705 GGACACCATTACAAACCAATCTGGAAGAGTTGTTTCATCTGCTCAACTTTCTACCCCGG 2764
Qy 1412 GGAAGTTTGGAAAGTTTGGAGGAGTTCCAGGAGGAGTTCCAAAGATATTAAATCAAGAGGAGC 1471
Db 2765 AGAGTTCCCAATTTGGAAGGTTTTTTGGAGGAGTTTGGAGGAGTTTGGACATTTGCCNAGAGGACC 2824
Qy 1472 AGATCTCAAGTTTGCACAAAATGTTGGCTCCACATTTGCTCAGAAAGGTAAAAAAGACG 1531
Db 2825 AGATAAAAAAATCGATGACATGCTGGGGCGCACATGTTGCGCGGCTCAAAAGCCGATG 2884
Qy 1532 TAATGAAGACATGCCCCCCHAAAAGAGCTCATTTTGGCGTTGATCTGAGCAGTCTGC 1591
Db 2885 TGTTCAAGAACATGCCCCCTCAAGACAGAACTAAATGTTGCGTGGAGCTGAGCCCTATGC 2944
Qy 1592 AGAAAGAAATATTACAAAGCTATTTTACCCTGAATTTATCAAGTATTGA---CAAAAAAGG 1648
Db 2945 AGAAGAAATACTACAGTACATCTCTCACTCGAAATTTTGAAGCACTCAATGCCCGAGGTG 3004
Qy 1649 GAGGTGCTCAAAATTTCCCTTAATTAACATATTATGATGGAATTACGAAAAGTATGCTGCCATC 1708
Db 3005 GTGGCAACCAAGGTGTTCTGTCTGAATGTTGTTGATGATGTTTAAAGAGTGTGCAACCATC 3064
Qy 1709 CTTA-----TATGCTAGAGGTGTTGAGCCAGTATTTCACGACGCAATCAAGCTT 1759
Db 3065 CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCTCAAGATCCCTAAATGGCATGTATGATG 3124
Qy 1760 TCAAAACAATTTTGGAGTCTTGTGGAAGCTGCAACTTCTAGATAAAATGATGTTCAAAAC 1819
Db 3125 CGATGCCCTTAATCAGACATCTGGGAAATTTATGCTGCTGCAGAAATGCTCAAGAAC 3184
Qy 1820 TGAAGACCAAGGACACAGATCTCTAATATACACACAGTATTCAGCATATGCTGCACTTAC 1879
Db 3185 TTAAGGAGGTGGGCATCGTGTACTCATCTTTTCCAGATGACCAAGATGCTAGACCTGC 3244
Qy 1880 TTGAAGACTACTGTACCCATAGAAATGGCAGTACGAGCGAATTGATGGAAGGTTGGCG 1939
Db 3245 TAGAGGATTTCTTGGAAATGAAGGTTTATAATACGAACGCATCGATGGTGGATCACTG 3304
Qy 1940 GAGCTGACGGCAATACGATAGATCGGTTCAATGCCCCAAATTTCTAAACAGTTTGTGTT 1999
Db 3305 GGAACATCGCGCAAGAGGCCATTGACCCTTCAATGACACCGGGTCTCAGCAGTTCTGCT 3364
Qy 2000 TTTTGTCTCTCAAGAGCTGGTGGCTTAGGAATAAATTTTGCAACGGCTGATACAGTAA 2059
Db 3365 TCTTGTCTTCCACTCGAGCTGGGGCGCTTGGAAATCAATCTGGCCACTGCTGACACAGTTA 3424
Qy 2060 TCATTATGACAGTGTACGTGGAATCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
Db 3425 TTATCTATGACTGTACTGTGGAACCCCCCAATAATGACATTCAGGCGCTTTAGCAGAGCTCAC 3484
Qy 2120 GACTTGGCCAAACAAATTAAGGTGATGATTTATAGGCTCATAAACCGAGGCCACCAATTGAAG 2179
Db 3485 GGATTTGGCAAAATATAAAGGTATGATCTACCGGTTTGTACCGCTGCGTCAAGTGAGG 3544
Qy 2180 AAAGATGATGCAATTTGACTAAAAAGAAAATGGTTCTAGACATCTTGTGTTGG----- 2234
Db 3545 AGCGCATCAGCGAGGTGGCAAGAAAGAAAATGATGCTGACGCATCTAGTGTGGCGGCTG 3604

Qy 2235 -----GAAACTCAAAACACAAACATTAAATCAGGAAGAGTTAGATGATCATCATCAGGTATG 2290
Db 3605 GGCTGGGCTCCAAGACTGGATCTATGTCCAAACAGAGCTTGTGATATCTCTCAATTTG 3664
Qy 2291 GATCAAAGAGAGCTTTTTTGTAGTGAAGATGATGAAGCAGGAAAGTCTGGAAAATTCATT 2350
Db 3665 GCACTGAGGAACATTATTCAGGATGAAGCCACTGATGGAGGAGGAGACAACAAAGAGGGAG 3724
Qy 2351 ATGATGATGGGCTATAGACAAATTCGTTGA-----TCGTGATCTCGTGGAGGCAGAGG 2404
Db 3725 AAGATGACAGTGTATCCACTACGATGATAAGGCCATTGAACGGCTGCTAGACCGTAACC 3784
Qy 2405 AAGTCTCAGTGGATGATGAAGAGGAGATGGAATCTTTAAAGGCTTTCAAGGTGGCTAAAT 2464
Db 3785 AGATGAGACTGAAGACACAGAATTCGAGGCAATGAATGATATTTGAGCTCATTCAAAG 3844
Qy 2465 TTGAATATATAGATGAAAATGAGGACAGCAGCATTTAGAGGCACAGAGAGTCTGCTGTA 2524
Db 3845 TGCCCAAGTATGTGGTACGGGAAGAAGAAATGGGGAGGAGAGGAGGTAGAAACGGGAAA 3904
Qy 2525 GCAATCTTCAGCAGGCAATTTCTGATAGAGCAAGTTATTGGGAAGAGTTCGTTAAAGATA 2584
Db 3905 TCATTAAACAGGAAGAAGTGTGATCTCTGA---CTACTGGGAGAAATTCGTCGGGACC 3961
Qy 2585 AATTTGAGCTGCACCGGCTGAGGAGCTTAATGCTCTTGGAAAAGGAAGAGAGTCGCA 2644
Db 3962 ATTATGAGCAGCAGCAAGAAAGATCTAGCCGAAATCTGGSCAAAGGAAAAGAAATCCGTA 4021
Qy 2645 AGCAGTTGGTATCCATTGAAAG---AAGATGATCTTCTGTTGGTGGAGATGTGAGCTCTG 2701
Db 4022 AACAGGTCAACTACAAATGATGGCTCCAGGAGGACCGAGATTGGCAGGACGACAGTCCG 4081
Qy 2702 ATGAGATGAAGTTATGAGCTGAGTCAACAGATGGTGAAGCAGCAGGACAGAGGATTC 2761
Db 4082 ACAACAGTCCGATTTACTCAGTGGCTTCAGAGGAAGTGATGAAGACTTTTGATGAACGTT 4141
Qy 2762 AGACGGTTCAGCGCGCTACAGAAAGAAAGGTCGCGATATTTTGGAAACCAA----- 2812
Db 4142 CAGAGCTCCCGTAGGCCAGTCGTAAGGCGCTCGGATGATAAAGATAAGCCATTGC 4201
Qy 2813 CTCGTTGATGGAAGGTGAGGGAGATCTTTTCAGAGTACTTGGGTTCAACAGAGTCAAA 2872
Db 4202 CTCCTCTGTTGGCCGCTGTTGGTGGGAATATTGAAGTACTTGGTTTTAATGCTCGTCAGC 4261
Qy 2873 GGCCCATTTTGTACAGACTTTGATGAGGTATCG-----AGCTGGCAATTTTGATT 2923
Db 4262 GAAAAGCCTTTCTTAATGCAATTTATGCGATATGATGCGCATGCGCCTCAGGATGCTTTACTA 4321
Qy 2924 GGAAGGAGTTTGTTCCTCGCTTAAAGCAGAAAGACCTTTGAAGAAATAAATGAATATGAA 2983
Db 4322 CCAGTGGCTTGTGAAGACCTCGGAGGCAATCAGAGAAAGAGTTCAAGGCATATGCT 4381
Qy 2984 TACTCTTTTGAAGACATTTGCTGAAGAAATAGACAGAAATCTCCAACTTTTCAGATG 3043
Db 4382 CTCCTTTTCATCGGCATTTATGTGAGCCGGGCGAGATGGGGCTGAGACCTTTGCTGATG 4441
Qy 3044 GTGTGCCCAAGGAGGAGCTTAGAATAGAAATGTTCTAGTCAGAAATGCTCTCTGATAC 3103
Db 4442 GTGTCCCGCAGAGGCGCTGCTCGCCAGCATGTCTTTACTAGAAATGGTGTATGCTT 4501
Qy 3104 TAGTTAGGAGAGAGGTGAAATTTTGAAGATCATCCAGGG 3144
Db 4502 TGATTCGAAGAGGTTCAAGGATTTGAACATGTTAATGGG 4542

RESULT 2

US-09-249-181A-1
; Sequence 1, Application US/09249181A
; Patent No. 644679
; GENERAL INFORMATION:
; APPLICANT: Seelig, Hans Peter
; APPLICANT: Renz, Manfred

;; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN

;; FILE REFERENCE: 8484-0059-999

;; CURRENT APPLICATION NUMBER: US/09/249,181A

;; CURRENT FILING DATE: 1999-02-12

;; PRIOR APPLICATION NUMBER: US 08/913,832

;; PRIOR FILING DATE: 1998-01-12

;; PRIOR APPLICATION NUMBER: PCT/DE96/00444

;; PRIOR FILING DATE: 1996-03-08

;; NUMBER OF SEQ ID NOS: 2

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 1

;; LENGTH: 6328

;; TYPE: DNA

;; ORGANISM: Homo sapiens

;; FEATURE:

;; NAME/KEY: CDS

;; LOCATION: (1)...(5736)

;; US-09-249-181A-1

Query Match 10.1%; Score 421.4; DB 4; Length 6328;
Best Local Similarity 51.9%; Pred. No. 1.2e-111;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;

QY	818	TACATCCATACCAAGCTTGAGGACTTAATTTTTCGGTCTCTCGTGTCAAAACAGACGC	877
DB	2177	TGCACCCCTATCAAAATGGAGGCGCTGAATTTGGTTCGCTTCTCTCGGCTCAGGGCCTG	2236
QY	878	ATGTAATCTTGTCTGTGTAATGGGACTAGCAAGACAAATTCAAAGCATTTGCCCTTTTAG	937
DB	2237	ACACATCTTGGCTGATGAGTGGCCCTTGGGAAATCTGTACACAGACAGTCTTCTGT	2296
QY	938	CTTACATTTTGGAGGAG-----AACCTCATTCGCAATTTGGTAATTTGCTTCCTCTAFCGA	991
DB	2297	ATTCCCTTTACAGAGGGGTCAATTCCAAAGGCCCTTCTTAGTGAGGCGCCCTCTTCTA	2356
QY	992	CTCTGCGTAATCGGAGAGAGTTTGGCCATGCGGCCCCACACAGATCAAGCTGTTATGT	1051
DB	2357	CCATCATCAATCGGAGCGGAGTTTGAATTTGGTTCGCTCAGACATGATGTGCTAACCT	2416
QY	1052	ATTTTGGCACTGCGCAAGCTCGACAGTATTCAGAGAAACATGATTTTACTTATCGAAAG	1111
DB	2417	ATGTGGGTGACAGGACAGCGGTGCCATCATCCGAGAGATGATTTCTCTTTGAAGACA	2476
QY	1112	ATCAAAAAAGATCAAGAAAAAGAAATCTGGCAAAATTAAGTAGCGAAAGCAAGCAAAAA	1171
DB	2477	ATGCCAATTCGTGGTGGCAAGAGCCCTC-----CCGCATGAAGAAAGAGGCAT	2524
QY	1172	GAATCAAGTTTGATGCTCTCTCATCTGATGATGATGATGATGATGATGATGATGATGATG	1231
DB	2525	CTGTGAATTTCCATGCTGCTGATCATCTTATGATGATGATGATGATGATGATGATGATG	2584
QY	1232	TAAACCAATTAAGTGGGAGTGATGATTTGATGAAGGTTCATCGACTGAAATAAAGG	1291
DB	2585	TGGGCTCTATTGATTTGGGCTGCTCATCTGTTGATGAAGCCCATCGGCTGAAGAACATC	2644
QY	1292	ATTCAAAGCTGTTCTCTTCAATGACACAGTATTCAGTAACCAAGCTGATTTCTTGACAG	1351
DB	2645	AGTCTAAGTTCTTCGGGTATTGAATGTTTACTCACTCCAGCACAAGCTGTGCTGACTG	2704
QY	1352	GAACACCACTTCAGAACCACTTGTGATGAATCTTTTTCATGCTCATGCAATTTCTTGATCGG	1411
DB	2705	GGACACCATTTACAAAAAATCTGAAGAGTTGTTTCTCATCTGCTCAACTTCTCACCCCG	2764
QY	1412	GGAAAGTTTGGAGTTTGGAGAGTTTCCAGGAGGTTTCAAGATATTAAATCAAGAGGAGC	1471
DB	2765	AGAGGTTCCCAATTTGGAAGTTTGGAGGAGTTTGGTGAATTTGCAATTCGCAAGAGGACC	2824
QY	1472	AGATCTCAAGGTTGCAAAAAATTTGCTCCACATTTTGGCTCAGAGGGGTAAAAAAGACG	1531
DB	2825	AGATAAAAAAATTCATCATCATCTGGGCGGCACATGTTGCGCGGCTCANAGCCGATG	2884
QY	1532	TAAATGAAGACATGCCCGCCCAAAAAAGGAGCTCATTTTGGGTGTTGATCTGACGAGTCTGC	1591

DB	2885	TGTTCAAGAACATGCCCTCCAAGACAGAACTAATTTGTGCGTGTGGAGCTGAGCCCTATGC	2944
QY	1592	AGAAAGAAATTTACAAAGCTATTTTACCCGTAATTTATCAAGTATTTGA---CAAAGAGG	1648
DB	2945	AGAAGAAATACACAAAGTACATCTCTCACTCGAAATTTTGAAGCACTCAATGCCGAGGTG	3004
QY	1649	GAGGTCTCAAAATTTCCCTTAATAACATTTATGATGGAATTTACGAAAGTATGTCGCCATC	1708
DB	3005	GTGGCAACCAAGGTGCTCTGCTGAATGTTGGTATGATCTTTAAGAAAGTGTGCAACCATC	3064
QY	1709	CTTTA-----TATGCTAGAGGGTGTGAGCCAGTATTACACGACGCAAAATGAAGCTT	1759
DB	3065	CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCTTAAGATGCTATGCAATGATGATG	3124
QY	1760	TCAACAACTTTTGGAGTCTTGTGAAAGCTGCAACTTTCTAGATAAAATGATGGTCAAAAC	1819
DB	3125	GCAGTGCCCTAATCAGAGCATCTGGGAAATTTATGCTGTCGAGAAATGCTCAAGAAC	3184
QY	1820	TGAAAGACGAGGACACAGAGTCTTAATATACACAGTTCACAGTTCACGATATGCTGACCTTAC	1879
DB	3185	TTAAGAGGGTGGGCATCTGTACTCATCTTTTCCAGATGACCAAGATGCTAGACCTGC	3244
QY	1880	TTGAACACTACTGTACCCCATAGAAATGGCAGTACGAGCGAAATTTGATGAAAGGTTGGCG	1939
DB	3245	TAGAGGATTTCTTGGAAACATGAAGTTAATACGACATCGATGTTGGTAATCACTG	3304
QY	1940	GAGCTGAGCGGCAAAATPACGATAGATCGGTTCAATCCAAATAATTTCTAACAGTTTGT	1999
DB	3305	GGAAATCGGCAAGAGGCCATTGACCGCTTCAATGCACCGGTGCTCAGCAGTTCTGCT	3364
QY	2000	TTTGTCTCTCCACAGAGCTGGTGGCTTAGGAATAAATCTTGCACCGGCTGATACAGTAA	2059
DB	3365	TCCTGCTTCCATCGAGCTGGGGCTTTGGAATCAATCTGGCCACTGCTGCACAGTTA	3424
QY	2060	TCATTTATCAGAGTGAATCCATCTGATCTTCAAGCAATGCGTGTAGAGCTCATC	2119
DB	3425	TTATCTATGATCTGATGAAACCCCATTAATGACATTCAGGCTTTAGCAGAGCTCAC	3484
QY	2120	GACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGACCATTTGAAG	2179
DB	3485	GGATTGGCCAAATAAAGGTAATGATCTACCGTTTGTGACCCGCTGCTCAGTGGAGG	3544
QY	2180	AAAGGATGATGCAATTTGACTAAAGAAATTTGTTCTAGAGCATCTTGTGTTGG-----	2234
DB	3545	AGCGCATACGCGAGTGGCAAGAAATGATGCTGACGCACTCTAGTGGTGGCGCTG	3604
QY	2235	-----GAACTCAAAACACAAACATTAATCAGGAAGTTAGTACATCATCAGGTATG	2290
DB	3605	GGCTGGGCTCCAGATCGGATCTATGTCCTCAACACAGGAGCTTGATGATATCTCTCAATTTG	3664
QY	2291	GATCAAGGAGCTTTTGTGCTAGTGAAGATGATGAAGCAGGAAAGTCTGGAAAAATTCATT	2350
DB	3665	GCATCAGGAACTATTCAAGGATGAAGCCACTGATGGAGGAGGAGACAAAGAGGGAG	3724
QY	2351	ATGATGATGGGCTATAGACAAATTCCTTTGA-----TCGTGATCTGTTGGAGGAGAGG	2404
DB	3725	AAGATAGCAGTGTATCCACTACGATGAAGGCCATTGAACCGCTGCTAGACCGTAACC	3784
QY	2405	AAGTCTCAGTGGATGATGAAGGAGATGGAATCTTTAAGGCTTTCAAGGTGCTTAATT	2464
DB	3785	AGGATGAGACTGAAGACACAGAAATTCGAGGGGCAATGAATGAATTTTGAAGCTCAATCAAG	3844
QY	2465	TTGAATATATAGATAAAATGAGGACAGCAGCATTTAGAGGCACAGAGAGTCGCTGCTGAAA	2524
DB	3845	TGCCCCAGTATGTGTTACGGGAAGAAATGGGGAGGAGGAGGAGGTAGAACGGGAA	3904
QY	2525	GCAATCTTTACAGAGCAATTTCTGATGAGCAAGTTTATTTGGGAAGAGTTGTTTAAAGATA	2584
DB	3905	TCATTAACACAGGAAGAAAGTGTGATCTCTGA---CTACTGGGAGAAATTTGCTGGCGACC	3961
QY	2585	AATTTGAGTGCACAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGGAGAGAGTCCGCA	2644
DB	3962	ATTATGAGCAGCAGCAAGAGATCTAGCCCGAAATCTGGGCAAGGAAAAAGAAATCCGTA	4021

Qy	2645	AGCAGTTGGTATCCATTGAAG---	AGATGATCTTGTGTTTGGAAAGATGTGAGCTCTG	2701
Db	4022	AACAGGTCAACTAATGATGGCTCC	CAGGAGGACCGAGATTGGCAGGACGACCAGTCCG	4081
Qy	2702	ATGGAGATGAAGTTATGAAGCTCAG	TCAACAGATGTTGAAGCAGCAGGACAGGAGTTC	2761
Db	4082	ACAACCAAGTCCGATTTACTCAGT	GGCTTTCAGAGGAAGTGATGAAGACTTTGATGAACGTT	4141
Qy	2762	AGACGGGTGCACGGCCGTACAGAA	GAAGGTCGCGATAATTTGGAAACAA-----	2812
Db	4142	CAGAAGCTCCCCGTAGGCCCAGTC	TAAAGGCCTCGGGAATGATAAGATAAGCCATTGC	4201
Qy	2813	CTCCGTTGTAGGAAGTGAAGGGAG	ATCTTTTCAGAGTACTGGGTTTCAACACAGAGTCAAA	2872
Db	4202	CTCCTCTGTTGGCCCGTGTGGTGG	GGAATATTGAAGTACTTGGTTTTAAATGCTCGTCAGC	4261
Qy	2873	GGGCCATTTTGTACAGACTTTGAT	GAGGTATGG-----AGCTGGCAATTTTTCATT	2923
Db	4262	GAAAGCCTTTCTTAATGCAATTA	TGCGATGTGATGCCACCTCAGGATGCTTTTACTA	4321
Qy	2924	GGAAGGAGTTTGTTCCTCGCTTA	AAGCAGAGACCTTTGAAGAAATAAATGAATATGAA	2983
Db	4322	CCCAAGTGGCTTGTAAAGAGACTG	CGAGGCAAAATCAGAGAAAGAGTTCAAGGCATATGCT	4381
Qy	2984	TACTCTTTTGAAGCACATTTGCTGA	AGAAATAGACAGGAATTTCTCAAACCTTTTCAGATG	3043
Db	4382	CTCTTTTTCATCGGGCATTTATGT	CAGCCGGGGCAGATGGGGCTTGAGACCTTTGCTGATG	4441
Qy	3044	GTGTGCCCAAGGAAGGACTTAGAAT	AGAAGATGTTCTAGTCAGAAATTTGCTTCTGTGATAC	3103
Db	4442	GTGTCCCCCGAAGAGGCCCTGTCT	CGCAGCATGTCTTACTAGAAATTTGGTGTATGCTTT	4501
Qy	3104	TAGTTCCAGGAAGGTGAAATTTCT	GAAAGATATCCAGGG	3144
Db	4502	TGATTTCCGAAGAAGGTTTCAGG	AGTTTTCGAACATGTTAATGGG	4542

```

RESULT 3
US-09-158-707-1
; Sequence 1, Application US/09158707
; Patent No. 6500923
; GENERAL INFORMATION:
; APPLICANT: Seelig, Hans Peter
; APPLICANT: Renz, Manfred
; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
; FILE REFERENCE: 8484-0043-999
; CURRENT APPLICATION NUMBER: US/09/158, 707
; CURRENT FILING DATE: 1998-09-22
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 6328
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(5736)
US-09-158-707-1

```

	Query Match	10.1%	Score 421.4;	DB 4;	Length 6328;
	Best Local Similarity	51.9%;	Pred. No. 1.2e-111;		
	Matches 1236;	Conservative 0;	Mismatches 1076;	Indels 69;	Gaps 10;
Qy	818	TACATCATACCACCTTGAGGCACTTAATTTTTCGCGTCTCTGTCATAACAGAGCC	877		
Db	2177	TGCACCCCTATCAAATGAGGGCCGTGAATTTGGTTTCGCGTCTCTCGGCTCAGGSCACTG	2236		
Qy	878	ATGTAATCCTTGCTGATCAATGGACTAGGCAAGACAATTCAAAGCATTCGCCCTTTTAG	937		
Db	2237	ACACCATCTTGGCTGATGAGATGGGCCCTTGGGAAACATGTACAGACAGAGCTCTCTCTGT	2296		
Qy	938	CTTCACCTTTTGGAGG-----AACCTCATTTCCGCCATTTGGTAATGTCTCCTCATCGA	991		

[illegible]

Db 3365 TCTTGCTTTCCACTCGAGCTGGGGCCCTTGGAAATCAATCTGGCCACTGCTGACACAGTTA 3424
Qy 2060 TCATTATGACAGTACTGGAATCTCATCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
Db 3425 TTATCTATGACTCTGACTGGAACCCCATATGACATCTAGGCTTTAGCAGAGCTCACC 3484
Qy 2120 GACTTGCCCAACAAATAAGGTGATGTTTATAGGCTCATAAACCGAGGCCACATTGAAG 2179
Db 3485 GGATTGGGCAAAATAAAAGGTATGATCTACCGGTTTGGACCCGGGCTCAGTGGAGG 3544
Qy 2180 AAAGGATGATGCAATGACTAAAGAAATAGTCTTAGAGCATCTTGTTGTTGG----- 2234
Db 3545 AGCGCATCAGCAGGTGGCAAGAGAAATGATGCTGACGCATCTAGTGTGGGCGCTG 3604
Qy 2235 -----GAAACTCAAAACAAACAAATTAATCAGGAAGTTAGATGACATCATCAGGTATG 2290
Db 3605 GGCTGGGCTCCAAGACTGGATCTATGTCCAAACAGGAGCTTGATGATATCTCAAAATTTG 3664
Qy 2291 GATCAAGAGGACTTTTGTAGTGAAGATGATGAGCAGGAAGCTGGGAAATTCATT 2350
Db 3665 GCACTGAGGAATCTCAAGGATGAAGCCACTGATGGAGGAGGAGACAAACAAAGAGGAG 3724
Qy 2351 ATGATGATGGCTTATAGACAAATTTGCTTGA-----TCGTGATCTCGTCAGGCAGAGG 2404
Db 3725 AAGATGACAGTGTATCCACTACGATGATAGGCCATTTGAACGGCTGTAGACCGTAACC 3784
Qy 2405 AAGTCTCAGTGGATGATGAAGAGGAGAAATGATTTCTTAAAGGCTTTCAAGGTGGCTAATT 2464
Db 3785 AGGATGAGACTGAAGACACAGAAATTCAGGCGCATGAATGATATTTGAGCTCATTTCAAAG 3844
Qy 2465 TTGAATATATAGTAAATGAGGCAGCAGCATTTAGAGGCACAGAGTCTGCTGTGAAA 2524
Db 3845 TGGCCCATGATGTGTGACGGGAAGAAATGCGGGGAGGAGGAGGTAGAAACGGGAAA 3904
Qy 2525 GCAAAATCTCAGCAGGCAATTCGATAGACAAAGTATTTGGAAGAGCTTCTTAAAGATA 2584
Db 3905 TCATTAACAGGAAGAAGTGTGGATCTGCA-----CTACTGGGAGAAATTTGCTGGGCACC 3961
Qy 2585 AATTGAGCTGCACAGGCTGAGGAGCTTAATGCTCTTTGGAAAAGAGAGAGTGGCA 2644
Db 3962 ATTATGAGCAGCAAGAAGATCTAGCCCGAAATCTGGGCAAGGAAAGAAATCCGTA 4021
Qy 2645 AGCAGTTGGTATCAATTGAAG-----AAGATGATCTTGCTGTTGGAAAGATGTAGCTCTG 2701
Db 4022 AACAGGTCAACTACAAATGATGGCTCCAGGAGGACCGAGATTTGGCAGGACGACAGTCCG 4081
Qy 2702 ATGGAGATGAAGTTATGAAGCTGAGTCAACAGATGCTGAAGCAGCAGGACAGGAGTTC 2761
Db 4082 ACAACAGTCGATTAATCTAGTGGCTTCAGAGGAGGTGATGAAGCTTTGATGAACGTT 4141
Qy 2762 AGACGGTTCGACGGCCGTACAGAGAAAGGGTCCGCAATAATTTGGAAACCAA----- 2812
Db 4142 CAGAACTCCCGTAGGCCAGTCTGAAGGCTCGCGAATGATAAGATGAAGCATTCG 4201
Qy 2813 CTCGTTGATGAAGGTGAGGGAGATCTTTTCAGAGTACTGGGTTTCAACAGAGTCAAA 2872
Db 4202 CTCCTCTGTGGCCGCTGTGTGGTGGGAATATGAAGTACTTGGTTTAAATGCTCGTCAGC 4261
Qy 2873 GGGCCATTTTGTACAGACTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT 2923
Db 4262 GAAAGCCCTTTCTTAATGCAATATGCGATATGATGATGCGACCTCAGGATGCTTTTACTA 4321
Qy 2924 GGAAGGAGTTTGTTCCTCGCTTAAAGCAGAGAGCTTTTGAAGAAATAAATGAATATGAA 2983
Db 4322 CCCAGTGGCTTGAAGAGACTCGAGGCAAAATCAGAGAAAGAGTTCAAGGCATATGCT 4381
Qy 2984 TACTCTTCTTGAAGCAATTTGCTGAAGAAATGACAGAGAAATTTCTCAACCTTTTTCAGATG 3043
Db 4382 CTCCTTTTCATCGCGCAATTTATGTGAGCCGGGGCAGATGGGCTTGAGACCTTTTCTGATG 4441
Qy 3044 GTGTGCCCAAGAGGAGCTTAGAATAGAGATGTTCTAGTCAGAAATGCTCTTCTGATAC 3103
Db 4442 GTGTCCCCCGAAGAGGCTGTCTCGCAGCATGTCTTACTAGAAATTTGGTGTATGTCTT 4501

Qy 3104 TAGTTCAGAGAGAGGTGAATTTGTAGAAGATCATCCACGG 3144
Db 4502 TGATTCGCAAGAGGTTCAGGAGTTTGAACATGTTAATGGG 4542

RESULT 4

US-09-620-312D-325
; Sequence 325, Application US/09620312D
; Patent No. 6569662

GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghaast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt.PL_genes Version 1.0
; SEQ ID NO 325
; LENGTH: 6475
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (219)..(5957)
US-09-620-312D-325

Query Match 10.1%; Score 421.4; DB 4; Length 6475;
Best Local Similarity 51.9%; Pred. No. 1.2e-111;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;

Qy 818 TATCATCATACAGCTTTGAGGAGCTTAAATTTTGGGTTCTCGTGGTCAAAACACAGCGC 877
Db 2395 TGCACCCCTATCAATGGAGGGCTGAATGGTTGGCTTCTCTGGGCTCAGGCACTG 2454
Qy 878 ATGTAATCTCTGCTGATGAATGGGACTAGGCAAGCAATTCAAAGCATTTGCCCTTTTAG 937
Db 2455 ACACCATCTTGGCTGATGAGATGGGCTTTGGGAAACTGTACAGACAGCAGTCTTCTGT 2514
Qy 938 CTTCACTTTTGGAGGAG-----AACCTCATTCGGCATTTGGTAATGTCCTCTATCGA 991
Db 2515 ATTCCCTTTTACAGGAGGGTCAATTCCAAAGGCCCTTCTCTAGTAGCGGCCCTCTTTCTA 2574
Qy 992 CTCTGCGTAACCTGGGAGAGAGATTTGCCACATGGSCCCACAGATGAACGTGTTATGT 1051
Db 2575 CCATCATCACTGGGAGCGGAGTTTGAATATGGGCTCCAGACATGTATGTCGTAACCT 2634
Qy 1052 ATTTTGGCACTGCGCAAGCTTCGAGCAGTATTATCAGAGAAATGAGTTTTACTTATCGAAAG 1111
Db 2635 ATGTGGGTGCAAGGACAGCCGTGCCATCATCCGAGAGATGAGTTCTCTCTTTGAAGACA 2694
Qy 1112 ATCAAAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGGAAAGCAAGCAAAAA 1171

Db 2695 ATGCCATCTGTTGGTGGCAAGAGCCCTC-----CCGCATGAAGAGGCAT 2742
 Qy 1172 GAATCAAGTTTGTATGCTCTCTCATCGTATGATGATCAACCTAGATTCAGCAGTTC 1231
 Db 2743 CTGTGAATTTCCATGTGTCTGATGATCTATGATGATGATGATGATGATGATGATGAT 2802
 Qy 1232 TAAACCAATTAAGTGGGAGTGCATGATGTTGATGAAGTGCATCGACTGAAAATTAAGG 1291
 Db 2803 TGGGCTCTATTGATTGGGCTTGCCTCATCGTGAATGAAGCCCATCGGCTGAAGAACATC 2862
 Qy 1292 ATTCAAAGCTTCTCTCTCATTTGACACAGTATTCAGATGAACACCGTATTCCTCTGACAG 1351
 Db 2863 AGTCTAAGTCTCTCGGGTATTGATGTTACTCACTCCAGCACAAAGCTGTGCTGACTG 2922
 Qy 1352 GAACACCACTTCAGAACAACTTTGATGAACATTTTTCATGCTCATGCTATTTCTTGATCGGG 1411
 Db 2923 GGACACCAATTACAAAACAACTGGAAGAGTGTGTTTCATCTGCTCAACTTCTCACCCCGG 2982
 Qy 1412 GGAAAGTTTGGAGTTTGGAGGAGTTCCAGGAGGAGTTCAAGATATTAAATCAAGAGAGC 1471
 Db 2983 AGAGGTTCCCAATTGGAAGGTTTTTGGAGGAGTTTGGTGAATTCGCAATTCGCAAGAGGACC 3042
 Qy 1472 AGATCTCAAGGTTGCACAAAATGTTGGCTCCACATTTGCTCAGAAAGGTTAAAAAGACG 1531
 Db 3043 AGATAAAAAACTGCATGACATGCTGGGGCGCACATGTTGCGCGGCTCAAGCCGATG 3102
 Qy 1532 TAATGAAGACATGCCCCCAGAAAGGAGCTCATTTTGGTGTGATCTGAGCAGTCTGC 1591
 Db 3103 TGTTCAGAAACATGCCCTCCAAGACAGAACTAAATGTCGTGTCGTGAGCTGAGCCCTATGC 3162
 Qy 1592 AGAAGATATTACAAAGCTATTTTACCGTAAATATCAAGTATTGA---CAAAAAGG 1648
 Db 3163 AGAAGAAATCTCAAGTACATCTCTCACTCGAAATTTTGAAGCACTCAATGCCCGAGGTG 3222
 Qy 1649 GAGTGTCTCAAAATTTCCCTTAATAACATTAATGATGGAATTAACGAAAGTATGCTGCCATC 1708
 Db 3223 GTGGCAACCGAGTGTCTCTGCTGATGTTGGTATGATGATCTTAAGAGTGTGCAACCATC 3282
 Qy 1709 CTTA-----TATGCTAGAGGTTGTAGCCAGTATTATCACAGCGCAATGAAGCTT 1759
 Db 3283 CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCTTAAGATGCTTAATGGCATGTATGATG 3342
 Qy 1760 TCBAACAACCTTTTGGAGTCTTGTGGAAGCTGCACTTCTAGATAAAATGATGTTCAAC 1819
 Db 3343 GCAGTCCCTTAATCAGAGCATCTGGGAAATTTTGTGCTGTCAGAAAATGCTCAAGAAC 3402
 Qy 1820 TGAAGAGCAAGGACACAGAGTCTCTAATATATACACACAGTTCACAGCATATGCTGACTTAC 1879
 Db 3403 TTAAGGAGGTGGGCATCGTCTACTCTCTTTCCAGATGACCAAGATGCTAGACCTGC 3462
 Qy 1880 TTGAAGACTACTGTACCCATGAAGAAATGGCGAGTACGAGCGAATTTGATGGAAGGTTGGCG 1939
 Db 3463 TAGAGGATTTCTTGGAACTATGAAGTTTAAATACGAACGATCGATCGTGGTGAATCACTG 3522
 Qy 1940 GAGCTGCGCGCAATACGCTAGATCGTTCAATGCCAAAAATTTCTAAACAAGTTTGT 1999
 Db 3523 GGAAATCGCGCAAGAGGCCAATGACCGCTTCAATGCAACCGGTTGCTCAGCAGTTCGCT 3582
 Qy 2000 TTTTGTCTCCACAAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGCTGATACAGTAA 2059
 Db 3583 TCTTGTCTTCCACTCGAGCTGGGGCTTGGAAATCAATCTGCGCCACTGCTGACACATTA 3642
 Qy 2060 TCATTTATGACAGTACTGGAATCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
 Db 3643 TTATCTATGACTCTGACTGGAACCCCAATATGATGATGATGATGATGATGATGATGATGAT 3702
 Qy 2120 GACTTGGCCAAACAAATGAAGTATGATTTATAGCTCATAAACGAGGACCATTTGAAG 2179
 Db 3703 GGATTTGGCAAAATAAAAGGTAATGATCTACCGGTTTGTACCCGTCGCTCAGTGAGG 3762
 Qy 2180 AAAGGATGATGCAATTTGACTAAAAAGAAATGGTTCTAGAGCATCTTGTGTTGG----- 2234

Db 3763 AGCGCATCACGAGTGGCAAGAAATGATGTCGACGCTATCTAGTGTGGCGCTG 3822
 Qy 2235 -----GAAACTCAAAACACAAACATTAATCAGGAAGTGTAGATGACATCATCAGGTATG 2290
 Db 3823 GGCTGGGCTCCAAAGCTGGATCTATGTCAAAACAGGAGCTTGTATGATATCTCTCAAAATTG 3882
 Qy 2291 GATCAAGAGCTTTTGTGCTAGTGAAGATGATGAAGCAGGAAAGTCTGGAATAATTCATT 2350
 Db 3883 GCACTGAGGAACTATTCAAGGATGAAGCCACTGATGGAGGAGGAGACACAAGAGGGAG 3942
 Qy 2351 ATGATGATCGGCTATPAGACAAATTCCTTGA-----TCGTGATCTCGTGGAGCGAGG 2404
 Db 3943 AAGATGACAGTGTATCCACTACGATGATAAGGCCATTTGAACGCTGCTAGACCGTAACC 4002
 Qy 2405 AAGTCTCAGTGGATGATGAAGAGGAATGGATTCTTAAAGGCTTTCAAGGTGGCTAATT 2464
 Db 4003 AGGATGAGACTGAAGACACAGAAATTCGAGGGCATGAATGAATATTGAGCTCATTTCAAAG 4062
 Qy 2465 TTCAATATATAGATGAAATAGGCGAGCAGCAATTAGAGGCACAGAGATCGCTGTCTGAA 2524
 Db 4063 TGCCCACTATGTTGTTACGGGAAGAATAATGGGGAGGAGAGGAGGTAGAACGGGAAA 4122
 Qy 2525 GCAAAATCTTACAGCAGCAATTTCTGATPAGACAAATTTATTGGGAAGAGTTGTTTAAAGATA 2584
 Db 4123 TCATTTAAACAGGAAGAAAGTGTGATCTCTGA---CTACTGGGAGAAATTTGCTCGGCACC 4179
 Qy 2585 AATTTGAGCTGCACAGGCTGAGGAGCTTAATGCTCTTGGAAAAAGAGAGAGTTCGCA 2644
 Db 4180 ATTATGAGCAGCAGCAGGAAGATCTAGCCGAAATCTGGGCAAGGAAAGAAATCCGTA 4239
 Qy 2645 AGCAGTTGGTATCCATTGAAG---AAGATGATCTTCTGTTTGGAGATGTGAGCTCTG 2701
 Db 4240 AACAGTCAACTACATGATGCTCCAGGAGACCGAGATTTGGCAGGACGACAGTCCG 4299
 Qy 2702 ATCGAGATGAAGTTTATGAAGCTGAGTCAACAGATGTTGAAGCAGCAGGAGAGTTC 2761
 Db 4300 ACAACAGTCCGATTTACTCAGTGGCTTCAAGGAAGTGTATGAAGACTTTGATGAACGTT 4359
 Qy 2762 AGACGGTGCAGCGGCTACAGAAAGGTCGCTGCGATTTTGGAAACCA----- 2812
 Db 4360 CAGAAGCTCCCGTAGGCCCGAGTCGTAAGGGCTCGGAAATGATAAGATTAAGCCATTGC 4419
 Qy 2813 CTCCTGTTGATGAAGTGAAGGAGATCTTTCAGAGTACTTGGGTTTCAACAGAGTCAAA 2872
 Db 4420 CTCCTCTGTTGGCCGCTGTTGGTGGAAATATGAAGTACTTGGTTTAACTGCTCGCAGC 4479
 Qy 2873 GGCCCAATTTTGTACAGACTTTGATGAGGTATGG-----AGCTGGCAATTTTGATT 2923
 Db 4480 GAAAAGCTTTCTTAATGCAATTTATGCGATATGCTATGCCACCTCAGGATGCTTTTACTA 4539
 Qy 2924 GGAAGGATTTGTTCTCGCTTAAAGCAGAGAGCTTTGAAGAAATAAATGAATATGGAA 2983
 Db 4540 CCAGTGGCTTGAAGAGACCTCGAGGCAAAATCAGAGAAAGAGTTCAAGGCATATGTCT 4599
 Qy 2984 TACTCTTCTTGAAGACATTTGCTGAAGAAATAGACGAGAATCTCAACCTTTTTCAGATG 3043
 Db 4600 CTCCTTTTCATCGCGATTTATGATGACCGGGGCGAGATGGGCTGAGACCTTTGCTGATG 4659
 Qy 3044 GTGTGCCCAAGGAGGACTTTAGAAATAGAAGATGTTCTAGTCAAGATTTGCTTCTGATAC 3103
 Db 4660 GTGTCCCGGAGAGGCTGTCTCGCCAGCATGCTCTTACTAGAAATGGTGTATGTCTT 4719
 Qy 3104 TAGTTCAAGGAGGAGGAAATTTGTAGAAGATCATCCAGG 3144
 Db 4720 TGATTCGCAAGAGGTTTCAAGGATTTGAACATGTTAATGGG 4760

RESULT 5
 US-09-376-594-75
 ; Sequence 75, Application US/09976594
 ; Patent No. 6673549
 ; GENERAL INFORMATION:
 ; APPLICANT: Furness, Michael

```

; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 75
; LENGTH: 5837
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 3125723CB1
US-09-976-594-75

```

Query Match	6.8%;	Score 285.4;	DB 4;	Length 5837;
Best Local Similarity	53.5%;	Pred. No. 4.3e-72;		
Matches 759;	Conservative 0;	Mismatches 591;	Indels 69;	Gaps 5;
QY	826	TACAGCTTCAGGGACCTTAATTTTTCGCGTCTCTCGTGGTCAAAACAGACGCATGTAATC	885	
DB	1584	TATCAGCTAGAGGTCTAAACTGCTAGCTCATCTCTGGTGCAAAATAATATAGTGTATC	1643	
QY	886	CTTGCTGATGAAATGGGACTAGGCAAGACAATTCAAAGCATGTGCCCTTTTAGCTTCACTT	945	
DB	1644	CTTGCTGATGAAATGGGCTTAGGAAAGACCATCCAGACCATATCATCTCTCTCTACCTG	1703	
QY	946	TTTGAGGAGAACCTCAT-----TCCGCATTTGGTAAATTGCTCTCTATCGACTCTGCGT	999	
DB	1704	TTCCACCAACACACAGCTGTATGCGCCCTTTCTTATAGTCGTCCCTTTATCCACCCCTCAC	1763	
QY	1000	AAC TGGGAGAGAGAGTTTGCCACATGGGCCCCCAAGATGAACGTGGTATATGTTATTTGGC	1059	
DB	1764	TCATGGCAGAGAGAGTTTGAATCTGGCACACAGATTAACGTAGTGGTTTACATAGG-	1822	
QY	1060	ACTGCGCAAGCTCGAGCAGTTATACAGAGACATGAGTTTTACTTATCGAAAGATCAAAA	1119	
DB	1823	-----TGACCTG	1829	
QY	1120	AAGATCAAGAAAAGAAATCTGGACAATAAGTAGCGAAAGCAAGCAAAAAGATCAAG	1179	
DB	1830	ATGAGCAGAAATACGATACGGGAATATGAATGGATTCAITCCCAACCAAAAAGATTGAAG	1889	
QY	1180	TTTGATGTCCTCTCACATCGATGAGATGATCAACCTAGATTACAGCAGTTCTTAAACCA	1239	
DB	1890	TTCAACGCACCTTATAACAACATATGAGATCCTCTTGAAAGATAAGACTGTGCTGGGCAGT	1949	
QY	1240	ATTAAAGTGGAGTGCATGATTGTTGATGAAGGTCTAGCTGAATAATAGGATTCAAAG	1299	
DB	1950	ATTAACTGGGCCCTTTCTGGGAGTGGATGAAGCCCATCGGTTGAAGATGATGACTCTTTA	2009	
QY	1300	CTGTCTCTCTTTCATTGACACAGTATTCAAGTAAACCAACCGTATCTCTTGACAGAAACCA	1359	
DB	2010	TTGTATAAACTCTGATTGATTTCAAGTCCAAACATAGGCTCTGTATTACGGGACCCCT	2069	
QY	1360	CTTCAGAACAACTTGGATGAACCTTTTCATGCTCATGCAATTTCTTGATGCGGGAGTGT	1419	
DB	2070	CTTCAGAAATTCCTCAAGAGCTCTGGTCTCTTGCTGCACTTATTATGCGGAGAAAGTTT	2129	
QY	1420	GGAAGTTTGGAGGATTCAGGAGGAGTTCAAAGATTAATTAAGAGGAGCAGATCTCA	1479	
DB	2130	GAATTTTGGGAAGATTTTGAAGAAGA---CCATGGGAAGGGAGAGAGATGGCTACGAG	2186	
QY	1480	AGGTTGCACAAAATGTTGGGCTCCACATTTGCTCAGAAGGGTAAAAAAGACGTAATGAAA	1539	
DB	2187	AGTCTTCATAGGTGCTAGAGCCTTTCTCTCGGAGATCAAAAAGATGTGGAGAA	2246	
QY	1540	GACATGCCCCCAAAAAGGAGCTCATTTTTCGTTGTTGATCTGAGCAGCTGTGCAAGAAAG	1599	
DB	2247	TCCCTTCTGTTAAAGTGAACAGATTCTCAGGTTGGAGATGTGAGGCCCTTCAGAAACAG	2306	

Qy	1600	TATTACAAAGCTATTTTTTACCCGTAATATATCAAGTATTTGACAAA---AAAGGAGAGTGCT	1655
Db	2307	TATTACAAAGTGGATCTTGACAGGAATTTACAGAGCTCTTTGCCAAGAGAAACAAGAGCGACG	2366
Qy	1657	CAAAATTCCTCTTAATAACATTTATGATGGAAATTACGAAAGATGCTGCCATCTTATATG	1716
Db	2367	ACATCTGGTTTTCTTAATATTTGATGGAACTGAAAAAATGTTGCAACCACTGCTATCTG	2426
Qy	1717	CTAGAGGGTGTTCAGCCAGTT---ATTCCAGCGCGCAAAATGAAAGCTTTTCAAAACAACATTTTG	1773
Db	2427	ATTAAACCCCTCGAAGAAATGAAAGGAAATGGACAGGAGATTTCTTGTCCTCCATCA	2486
Qy	1774	GAGTCTTTGTGGAAAGCTGCAACTTCTAGATAAAATGATGGTCAAACTGAAAGAGCAAGGA	1833
Db	2487	AGGAGCAGTGGGAAGTTGATTTTATTAGACAAACTGTTCACAAGACTTTCGAGAAAGGGGG	2546
Qy	1834	CACAGAGTCTAATATACACACAGTTTTCAGCATATGCTGGACTTACTTGAAGACTACTGT	1893
Db	2547	AATCGAGTGTATATCTTCTCTCAGATGGTGAGAAATGTTGGATATCTTGGCTGGAATACCTTA	2606
Qy	1894	ACCCATAAGAAATGGCAGTACGAGCGCAATTTGATGGAAAGGTTGGCGAGCTGAGCGGCAA	1953
Db	2607	ACTATTAAACACTATCTCTTTCAGCGCTCGGATGGTTCCATCAAGGGAGAAATCCGAAAA	2666
Qy	1954	ATACGCATAGATCGGTTCAATGCGCAAAAAATTTCAACAAGTTTTTTTTTGTCTCTCCACA	2013
Db	2667	CAGSCACTGCACACTTCAATGCAGATGGTCTCGAGGACTTCTGTTTCTGCTCTCGACA	2726
Qy	2014	AGAGCTGGTGGCTTAGGAATAAATCTTGCNACCGCTGATACAGTAATCATTTTATGACAGT	2073
Db	2727	AGGGCTGTGTGGCTCGGGAATCAATTTGGCTTCAGCGGACACAGTCGTCTCTTTGACTCT	2786
Qy	2074	GACTGGAATCCTCATGCTGATCTTTCAAGCAATCGCTAGAGCTCATCGACTTGCGCCAAAA	2133
Db	2787	GACTGGAAACCCAGAAATGACTTGCAGGCACAAAGCCGAGCGCATAGAAATGGTCAGAAG	2846
Qy	2134	AATAAGGTGATGATTTATAGGCTCATAAACCGAGGCCACCATTGAAGAAAAGGATGATGCAA	2193
Db	2847	AAGCAGGTAATATTTTACCGCTTAGTTTACAAAGGGGACTGTGGAGGAGGAGATCATAGAA	2906
Qy	2194	TTGACTAAAAAGAAAATGGTTCTTAGAGCATCTTTGTTGTT	2232
Db	2907	CGGCGCCAAAAGAAAGATGGTATTAGATCATCTGCTGATT	2945

RESULT 6

US-09-702-705-1802

; Sequence 1802, Application US/09702705

; Patent No. 6504010

; GENERAL INFORMATION:

; APPLICANT: Wang, Tongtong

; APPLICANT: Bangur, Chaitanya S.

; APPLICANT: Lodes, Michael A.

; APPLICANT: Fanger, Gary

; APPLICANT: Vedvick, Tom

; APPLICANT: Carter, Darrick

; APPLICANT: Retter, Marc

; APPLICANT: Mannion, Jane

; APPLICANT: Fan, Liqun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.478C14

; CURRENT APPLICATION NUMBER: US/09/702,705

; CURRENT FILING DATE: 2000-10-30

; NUMBER OF SEQ ID NOS: 1833

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1802

; LENGTH: 2512

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-09-702-705-1802

RESULT 6

```

US-09-702-705-1802
; Sequence 1802, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1802
; LENGTH: 2512
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-702-705-1802

```



```
Db 1047 ACTGAGCTAGTGGGAAGCTTACCTGCTGGATPAAGCTACTAGCATTCCTGATTCCTGGG 1106
Qy 1831 GGACACAGAGTCCTAATATACACACAGTTTTCAGCATATGCTGACCTTACTTTGAAGACTAC 1890
Db 1107 GGCCATCGGTTTACTTTTCTCCAAATGACCAGATGTTGATATTTCTCCAAGACTAT 1166
Qy 1891 TGTACCCATAGAATAAGGAGTAGGAGAGCAATTTGATGAAGAGTTGGCGAGCTGAGCGG 1950
Db 1167 ATGGATTACAGAGCTACAGCTATGAGCGTGTGGATGTTCTGTGAGAGGAGAAGAGA 1226
Qy 1951 CAAATACGATAGATCGTTCAATGCCCCAAATTTCTAAAGTTTCTTTTCTCTCC 2010
Db 1227 CACTTGGCCATTA-----GAACCTTGGACAGAGCCCAATTTCTGTTTCTCTGAGT 1280
Qy 2011 ACAAGAGCTGGTGGCTTAGGAATAAATCTTGGCAAGCGCTGATACAGTAATCATTTATGAC 2070
Db 1281 ACTAGGSCAGGTGGAGTTGGCATGAATTAACAGCAGCAGATCTGTGATTTTGTGAC 1340
Qy 2071 AGTACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATCGACTTGGCCAA 2130
Db 1341 AGTGACTTTAATCCTCAGAATGACTTGAAGCAGCTGCCAGGCTCATCGCATTTGGCCAA 1400
Qy 2131 ACAATAAGGTGATGATTTATAGCTCATAAACGAGGCGACCATTTGAAGAAA 2182
Db 1401 ACAAAGTCTGTTAAAGTTATTCGGCTGATTGGTCGAGACACTGTGGAAGAAA 1452
```

RESULT 8

```
US-09-671-325-1802
; Sequence 1802, Application US/09671325
; Patent No. 6667154
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Panger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C12
; CURRENT APPLICATION NUMBER: US/09/671,325
; CURRENT FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 1825
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1802
; LENGTH: 2512
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-671-325-1802
```

```
Query Match 5.1%; Score 213.6; DB 4; Length 2512;
Best Local Similarity 53.8%; Pred. No. 1.9e-51;
Matches 544; Conservative 0; Mismatches 444; Indels 24; Gaps 4;

Qy 1180 TTTGATGCTCTCCATCGATGATGATGATCAACCTAGATTCAGCAGTTCCTAAACCA 1239
Db 456 TTTGATGCTCTCCATCGATGATGATGATCAACCTAGATTCAGCAGTTCCTAAACCA 515
Qy 1240 ATTAAGTGGGAGTGCATGATTTGATGAAGGTTCATCGCTGAAATAAGGATTCAAAG 1299
Db 516 TTCCCTTGGAGTGTCTTGTGTGGATGAAGCTCAGAGTTGAAACCAAGCTCCCTG 575
Qy 1300 CTGTCTCTTTCAATGACACAGTATTCAGTAACACCGTATTTCTTGCAGGAGAACCA 1359
Db 576 CTGATAAGACCTTGTGAGAGTCTCAGTAGTCTTCAAGTCTCTGTTGACCGGAATCCC 635
Qy 1360 CTTGAGAACAACTGGATGAACCTTTTCATGCTCATGATTTTCTTGTGCGGGAAGTTT 1419
```

```
Db 636 ATCCAGAACAGCCTCCAAGAGCTCTACTCCCTCCTCAGTTTCTGGAGCCTGATCTCTTT 695
Qy 1420 -----GGAGTTTGGAGAGTTCAGAGAGAGTTTCAAAGATATTAATCAAGAGAGAGAG 1473
Db 696 TCCAAGGAAGAGTGGAGATTTTATTTCAACGCTACAGGATATTGAGAAAGAAATCTGAG 755
Qy 1474 ATCTCAAG--GTTGCACAAATGTTGGCTCCACATTTGCTCAGAAGGGTAAAAAAGAC 1530
Db 756 TCAGCAAGTGAACTGACAAACTCTTGCAGCCATTTCTGCTGAGGAGTGAAGAGCTGAG 815
Qy 1531 GTAATGAAAGACATGCCCCCAAAAAGAGAGCTCATTTTCCGTTGTTGATCTGAGCAGCTCTG 1590
Db 816 GTAGCTACAGAGCTTCCCAAGGAAGACAGAGTAGTATACCATGGCATGTGAGCATTTG 875
Qy 1591 CAGAAAGAAATATTAACAAGCTATTTTACCCTGTAATATTAAGTATTTGACAAAAAAGGGA 1650
Db 876 CAGAAAGAAATACTACAAGGCCATTTTGTATGAAGAGCTAGATGCAATTTGAAATATGAGACG 935
Qy 1651 GGTGCTCAAAATTTCCCTTAATTAACATTTATGATGGAATTAAGAAAGTATGCTGCCATCCT 1710
Db 936 GCAAGAAAGGTTAAACTACAGAACATTTTGTCCAGCTTTCGAAAGTGTGTGGATCACCCA 995
Qy 1711 TATATCTAGAGGCTTTCAGCCAGTTTATTCACGACGCAAAATGAAGCTTTTCAAAACAATTT 1770
Db 996 TATTTGTTGATGTTGTGGAGCC-----GGAGCCTTTTGAAGTTGGAGACCACTG 1046
Qy 1771 TTGGAGCTTTGTGGAAGCTGCAACTTTCTAGATAAAATGATGGTCAAACTGAAGAGCAA 1830
Db 1047 ACTGAGGCTAGTGGGAAGCTTCACTGCTGGATAAGCTACTAGCATTCCTGTATTCTGGG 1106
Qy 1831 GGACACAGAGTCTTAATATACACAGTTTACAGCATATGCTGGACTTACTTGAAGACTAC 1890
Db 1107 GGCCATCGGCTTTTACTTTTCTCCCAATGAAGCTGATGATGATTTCTTCCAGACTAT 1166
Qy 1891 TGTACCCATAAGAAATGGCAGTACGAGCGAAATTTGATGGAAGGTTGCGCGAGCTGAGCGG 1950
Db 1167 ATGGATTACAGAGCTTACAGCTATGAGCGTGTGGATGTTCTGTGAGAGAGAGAGA 1226
Qy 1951 CAAATACGATAGATCGGTTCAATGTCGCAAAATTTAAACAAGTTTGTGTTTGTCTCTCC 2010
Db 1227 CACTTGGCCATTA-----GAACTTTGACAGCAGCAGCCCATTTTCTCTCTGAGT 1280
Qy 2011 ACAAGAGCTGGTGGCTTAGGAATAAATCTTGAACGCTGATACAGTAAATCATTTATGAC 2070
Db 1281 ACTAGGCGAGGTGGAGTTGGCATGAATTAACAGCAGCAGACTACTGTGATTTTGTGAC 1340
Qy 2071 AGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATCGACTTGGCCAA 2130
Db 1341 AGTGACTTTAATCTCAGATGATGATTTGCAAGCAGCTGCCAGGCTCATCGCATTTGGCCAA 1400
Qy 2131 ACAATAAGGTGATGATTTTATAGGCTCATAAACCGAGGCGACCATTTGAAGAAA 2182
Db 1401 ACAAAGTCTGTTAAAGTTATTCGGCTGATTGGTCGAGACACTGTGGAAGAAA 1452
```

RESULT 9

```
US-09-579-181-11
; Sequence 11, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chrivia, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 11
; LENGTH: 8916
; TYPE: DNA
```

; ORGANISM: Human
US-09-579-181-11

Query Match 2.7%; Score 112.2; DB 4; Length 8916;
Best Local Similarity 53.4%; Pred. No. 1.4e-21;
Matches 259; Conservative 0; Mismatches 223; Indels 3; Gaps 1;

QY 1759 TTCAAAACAACTTTTGGAGTCTTTGGAAAGCTGCAACTTCTAGATAAAATGATGTCAAA 1818
DB 5326 TTAGACTCTCCAGTATGTTGGGAAGTTGAGACGTTGGCAGTGTGTTGGCGCAG 5385
QY 1819 CTGAAGAGCAAGACACAGAGTCTTAATATACACAGTTTTCAGCATATGCTGGACTTA 1878
DB 5386 CTCAAGGAGAGGCGCACCGAGTCTCATCTTCAACCCAGATGACCCGAATGCTGGATGTA 5445
QY 1879 CTTGAAGACTACTTACCCATAGAAATGGCAGTACGAGCGAATTTGATGGAAGTTGGC 1938
DB 5446 TTGGAGCAGTTTCTCACTACCATGCGCATCTCTACCTGCGCTGGATGATCTACTAGA 5505
QY 1939 GGAGCTGAGCGCAAAATACGATAGATCGGTTCAATGCCAAAATTTCTAACAAAGTTTGT 1998
DB 5506 GTTGAACAGACAGAGCGCTTTCATGGAAGGTTCAATGCAGACAA---ACGCATATTCTGC 5562
QY 1999 TTTTGTCTTCCCAAGAGCTGGTGGCTTAGGAATAAATCTTGAACGGCTGTATACAGTA 2058
DB 5563 TTCACTCTTCACTCGAGTGGGGTGTGGCGGTGAACCTGACAGGAGCAGACACTGTT 5622
QY 2059 ATCATTTATGACAGTGTGGAATCTTCATGCTGATCTTCAAGCAATGGCTAGAGTCAT 2118
DB 5623 GTTTTATGACAGCGACTGGAATCCCAACCATGATGCTCAGGCCCGAGCCGCTGTCAC 5682
QY 2119 CGACTTGGCCAAACAAATAGGTGATGATTTATAGGCTCATAAACCGAGGCACCATGAA 2178
DB 5683 CGAATTTGGCCAGACCCGGGATGTCCACATATATAGGCTTATCAGTGAACGACAGTGGAG 5742
QY 2179 GAAAGGATGATGCAATTTGACTTAAAGAAAATGTTCTAGAGCATCTTGTGTTGGGAAA 2238
DB 5743 GAGAACATCTTAAAGGCAAAATCAGAGAGAATGTTGGGGGACATGGCCATTTAGGGA 5802
QY 2239 CTCAA 2243
DB 5803 GGCAA 5807

RESULT 10
US-09-579-181-10
; Sequence 10, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chivia, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 10
; LENGTH: 9354
; TYPE: DNA
; ORGANISM: Human
US-09-579-181-10

Query Match 2.7%; Score 112.2; DB 4; Length 9354;
Best Local Similarity 53.4%; Pred. No. 1.5e-21;
Matches 259; Conservative 0; Mismatches 223; Indels 3; Gaps 1;

QY 1759 TTCAAAACAACTTTTGGAGTCTTTGGAAAGCTGCAACTTCTAGATAAAATGATGTCAAA 1818
DB 5764 TTAAGACTCTCCAGTATGATTTGGGAAGTTGACAGAGTGTGGCAGTGTGTTGGCGCAG 5823

QY 1819 CTGAAGAGCAAGGACACAGAGTCTTAATATATACACAGTTTCAGCATATGCTGGACTTA 1878
DB 5824 CTCAAGGCGAGAGGGCCACCGAGTGTCTCATCTTCAACCCAGATGACCCGAATGCTGGATGTA 5883
QY 1879 CTTGAAGACTACTGTATACCCATAGAAATGGCAGTACGAGGAAATTTGATGGAAGGTTGGC 1938
DB 5884 TTGGAGCAGTTTCTCACTACCATGCGCATCTCTACCTGCGCTGGATGATCTACTAGA 5943
QY 1939 GGAGCTGAGCGCAAAATACGATAGATCGGTTCAATGCCAAAATTTCTAACAAAGTTTGT 1998
DB 5944 GTTGAACAGACAGAGCGCTTTCATGGAAGGTTTCATGTCAGACAA---ACGCATATTCTGC 6000
QY 1999 TTTTGTCTTCCCAAGAGCTGGTGGCTTAGGAATAAATCTTGAACGGCTGTATACAGTA 2058
DB 6001 TTCACTCTTCACTCGAGTGGGGTGTGGCGGTGAACCTGACAGGAGCAGACACTGTT 6060
QY 2059 ATCATTTATGACAGTGTGGAATCTTCATGCTGATCTTCAAGCAATGGCTAGAGTCAT 2118
DB 6061 GTTTTATGACAGCGACTGGAATCCCAACCATGATGCTCAGGCCCGAGCCGCTGTCAC 6120
QY 2119 CGACTTGGCCAAACAAATAGGTGATGATTTATAGGCTCATAAACCGAGGCACCATGAA 2178
DB 6121 CGAATTTGGCCAGACCCGGGATGTCCACATATATAGGCTTATCAGTGAACGACAGTGGAG 6180
QY 2179 GAAAGGATGATGCAATTTGACTTAAAGAAAATGTTCTAGAGCATCTTGTGTTGGGAAA 2238
DB 6181 GAGAACATCTTAAAGGCAAAATCAGAGAGAATGTTGGGGGACATGGCCATTTGAGGGA 6240
QY 2239 CTCAA 2243
DB 6241 GGCAA 6245

RESULT 11
US-09-023-655-429
; Sequence 429, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 429:
; SEQUENCE CHARACTERISTICS:

```
; LENGTH: 1868 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HEARFET01
; CLONE: 1645339
US-09-023-655-429

Query Match      2.5%; Score 103.8; DB 4; Length 1868;
Best Local Similarity 65.1%; Pred. No. 1.4e-19;
Matches 153; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 1988 ACAAGTTTGTCTCTCCACAGAGCTGGTGGCTTAGGAATAAATCTTGGCAACGG 2047
Db 55 ATATCTTTGTGTTTCTGCTATCAACAAAGCTGGTGGATTAGGAATAAATCTGACTTCAG 114

QY 2048 CTGATACAGTAATCATTTATGACAGTGCAGTGGTAATCCTCATGCTGATCTTCAAGCAATGG 2107
Db 115 CAAATGTTGTATACCTTCAGATATTGACTGTAACTCTTATATGACAAACAGCAGAG 174

QY 2108 CTAGAGCTCATCGACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAG 2167
Db 175 ATAGATGCCATAGAGTAGGCCAGACTAAAGAGTACTAGTTATATAAATTAAGCCAG 234

QY 2168 GCACCATTTGAAGAAGATGATGCAATTTGACTAAAGAAATGGTTCTTAGAGCA 2222
Db 235 GGACGATTGAAGATCCATCTATAAATAACCAACAGAAATTCGAAACTAGAACCA 289

RESULT 12
US-09-328-111-553/c
; Sequence 553, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 553
; LENGTH: 575
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(575)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-553

Query Match      2.5%; Score 103.2; DB 3; Length 575;
Best Local Similarity 55.2%; Pred. No. 9.5e-20;
Matches 223; Conservative 0; Mismatches 178; Indels 3; Gaps 1;

QY 1241 TTAGTGGAGTGCATGATTTGATGAGGTGATCGATCGATGAAATTAAGGATTCGAAGC 1300
Db 419 TAAATGGCATTTATAGGTGTTGATGAAGCACCAGATTAAAGAAATGATGATCTCCCTTC 360
```

```
QY 1301 TGTTCCTCTTATGACACAGTATTCAAGTAACACCGTATTCTTCTGACAGGACACCCAC 1360
Db 359 TGTATAAAACCTTTAATAGATTTTAAATCCAATCATCGTCTCCTTATCATCTGGAACCTCC 300
QY 1361 TTCAAGAACAACTTTGGATGAACCTTTTCATGCTCATGCACTTTCTTGTATGCGGGGAAGTTTG 1420
Db 299 TACAGAAATTCCTCAAGAGAGCTCTGGTCTTTGCTACATTTTCATTATGCCAGAAAGTTT 240
QY 1421 GAAGTTTGAGAGTTCCAGGAGAGTTCAGAGATTAATCAAGATATTAATCAAGAGGAGCAGATCTCAA 1480
Db 239 CTTCTCGGGAAGATTTTGAAGAAGA---ACATGGCAAGGAGAGAGATATGTTTATGCAA 183
QY 1481 GGTTCGCAAAAATGTTGGCTCCACATTTGCTCAGAAAGGTAAAGAGAGCTAATGAAAG 1540
Db 182 GCCTTCACAAGAGAGCTTGAGCCATTTCTGTTACGCCGAGTTAAGAAAGATGTGCAAAAAT 123
QY 1541 ACATGCCCCCCCCAAAAGAGAGCTCATTTTCGCTGTTGATCTGAGCAGTCTGCAGAAAAGAT 1600
Db 122 CTCCTCCTCCCAAGGTTGAGCAGATTTTAAGAAATGGAATGAGTGTCTTTACAGAAACAT 63
QY 1601 ATTACAAAGCTATTTTACCCTTAATTAATCAAGTATTTGACAAA 1644
Db 62 ATTACAAATGGATTTTAACTAGAAATTACAAAGCCCTCAGCAA 19
```

```
RESULT 13
US-09-535-008-60
; Sequence 60, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60
; LENGTH: 5386
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (75)..(1805)
US-09-535-008-60
```

```
Query Match      2.4%; Score 100.2; DB 4; Length 5386;
Best Local Similarity 52.8%; Pred. No. 3.2e-18;
Matches 216; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 1778 CTTGTGGAAGCTGCAACTTCTAGATAAATGATGTCTAAACTGAAAGAGCAGGACACA 1837
Db 3222 CCTCGGGTAAATTTGAGCTTCTTGATAGAATTTCTCCCAAACTCCGAGCAACCAACCA 3281
QY 1838 GAGTCTTAATATACACAGATTTTCAGATATGCTGCACTTACTTGAAGACTACTGTACCC 1897
Db 3282 AAGTGTCTGCTGTTCTGCCAAATGACCTCCCTCATGACCATCATGGAAGATTTACTTTGCGT 3341
QY 1898 ATAAGAAATGGCAGTACGAGCGAATTTGATGGAAGGTTGCGGAGCTGAGCGGCAATAC 1957
Db 3342 ATCGCGGCTTTAAATACCTCAGGCTTGATGGAACACAGAGGCGGAGGACCGGGCATGC 3401
QY 1958 GCATAGATCGGTTCAATGCGCAAAAATTTCTAACAAGTTTTGTTTTTCTCTCCACAAGAG 2017
Db 3402 TGTGTAACACCTTCAACGAGCCGGCTCTGAGTACTTCTATCTCTGCTCAGCACCAGG 3461
QY 2018 CTGGTGGCTTAGGAATAAATCTTTGCAACGGCTGATACAGTAATCATTTATGACAGTACT 2077
```

Db 3462 CTGGGGGCTCGGCTGAACCTCCAGTCGGCAGACACTGTGATCATTTTTCAGACGACT 3521
Qy 2078 GGAATCCTCATGCTGATCTTCAAGCAATGGTAGAGCTCATCGACTTGGCCAAACAATA 2137
Db 3522 GGAATCCTCACCAGGACTGCAAGCGCAGGACCGACCGCCACCGCATCGGGCAGCAGAACG 3581
Qy 2138 AGGTGATGATTATAGGCTCTATAAACCGAGGACCAATTGAAGAAAGGAT 2186
Db 3582 AGGTGCGTGTCTCGGCTCTGCACCGTCAACAGCGTGGAGGAGAAGAT 3630

RESULT 14

US-09-535-008-66
; Sequence 66, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 66

; LENGTH: 5468
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (75)..(5012)
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (1780)
; OTHER INFORMATION: GenBank Accession No. 6465629 U29175 shows a C at this
; OTHER INFORMATION: position (position 1784 in GenBank) rather than
; OTHER INFORMATION: the G shown here.
; FEATURE:
; NAME/KEY: allele
; LOCATION: (26)
; OTHER INFORMATION: Polymorphism of either T or C in this noncoding
; OTHER INFORMATION: region.
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1583)
; OTHER INFORMATION: Polymorphism of A or G resulting in a silent
; OTHER INFORMATION: mutation.
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1598)
; OTHER INFORMATION: Polymorphism of T or C resulting in a silent
; OTHER INFORMATION: mutation.
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1892)
; OTHER INFORMATION: Polymorphism of A or G resulting in a silent
; OTHER INFORMATION: mutation.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4498)..(4499)
; OTHER INFORMATION: A CAG is missing between these bases as compared
; OTHER INFORMATION: to SEQ ID NO:1.
US-09-535-008-66

Query Match 2.4%; Score 100.2; DB 4; Length 5468;
Best Local Similarity 52.8%; Pred. No. 3.2e-18;
Matches 216; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

Qy 1778 CTTGTGGAAGAGCTGCAACTCTTAGATAAAATGATGTCAAACTGAAAGAGCAAGGACACA 1837
Db 3307 CCTCGGTAATTTTGAAGTCTTTGATAGAAATCTTCCCAAACCTCCGAGCAACCAACACA 3366
Qy 1838 GAGTCTTAATATACACACAGTTTCAGCATATGCTGGACTTACTTTGAAGACTACTGTACCC 1897
Db 3367 AAGTGTGCTGTTTCTGCCAAATGACCTCCCTCATGACCATCATGGAAGATTACTTTGCGT 3426
Qy 1898 ATAAGAAATGGCAGTAGCAGCAATTTGATGGAAGGTTGGCGGAGCTGAGCGCAATAATC 1957
Db 3427 ATCGCGGCTTTAAATACCTCAGGCTTGATGGAACCAAGAGCGGAGGACCGGGGATGC 3486
Qy 1958 GCATAGATCGGTTCAATGCAAAAATTTCTAAACAAGTTTGTGTTTTCCTCTCCACAAGAG 2017
Db 3487 TGCTGAAAACCTTCAACGAGCCCGGCTCTGAGTACTTTCATCTTCTGCTCAGCACCCGG 3546
Qy 2018 CTGGTGGCTTAGGAATAAATCTTGCACCGGTGATACAGTAATCATTTATGACAGTACT 2077
Db 3547 CTGGGGGCTCGGCTGAAACCTCCAGTCGGCAGACACTGTGATCATTTTGTGACAGGACT 3606
Qy 2078 GGAATCCTCATGCTGATCTTCAAGCAATGCTAGAGCTCATCGACTTGGCCAAACAATA 2137
Db 3607 GGAATCCTCACAGGACCTGCAAGCGCAGGACCGACCCACCGCATCGGCGCAGCAACG 3666
Qy 2138 AGGTGATGATTTATAGGCTCATAAACCGAGGACCAATTGAAGAAAGGAT 2186
Db 3667 AGGTGCTGTGCTCGGCTCTGCACCGTCAACAGCGTGGAGGAGAAGAT 3715

RESULT 15

US-09-535-008-1
; Sequence 1, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5471
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (75)..(5015)
; FEATURE:
; NAME/KEY: conflict
; LOCATION: (1780)
; OTHER INFORMATION: GenBank Accession No. 6465629 U29175 shows a C at this
; OTHER INFORMATION: position (position 1784 in GenBank) rather than
; OTHER INFORMATION: the G shown here.
; FEATURE:
; NAME/KEY: allele
; LOCATION: (26)
; OTHER INFORMATION: Polymorphism of either T or C in this noncoding
; OTHER INFORMATION: region.
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1583)
; OTHER INFORMATION: Polymorphism of A or G resulting in a silent
; OTHER INFORMATION: mutation.
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1598)
; OTHER INFORMATION: Polymorphism of T or C resulting in a silent
; OTHER INFORMATION: mutation.
; OTHER INFORMATION: Polymorphism of T or C resulting in a silent

; OTHER INFORMATION: mutation.

; FEATURE:

; NAME/KEY: allele

; LOCATION: (1892)

; OTHER INFORMATION: Polymorphism of A or G resulting in a silent

; OTHER INFORMATION: mutation.

US-09-535-008-1

```
Query Match      2.4%; Score 100.2; DB 4; Length 5471;
Best Local Similarity 52.8%; Pred. No. 3.2e-18;
Matches 216; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

Qy 1778 CTTGTGGAAGCTCCAACTCTAGATAAATGATGGTCAAACTCAAAAGACGAAGGACACA 1837
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3307 CCTCGGTAATTTGAGCTTCTTGATAGAATTTCTCCAAACTCCGAGCAACCAACCACA 3366
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1838 GAGTCCTAATATACACACAGTTTCAGCATATGCTGGACTTACTTGAAGACTACTGTACCC 1897
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3367 AAGTGCTGCTTCTGCCAAATGACCTCCCTCATGACCATCATGGAAGATTACTTTGCGT 3426
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1898 ATAAGAAATGGCAGTACGAGCGAATTGATGAAAGTTGGCGGAGCTGAGCGGCAATAC 1957
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3427 ATCGCGCTTTAAATACCTCAGGCTTGATGGAACCAAGCGGAGGACCGGGGCATGC 3486
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1958 GCATAGATCGGTTCAATGCCAAAAATTTCTAACAGTTTTTGTCTTCTCCACAAGAG 2017
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3487 TGTGAAAAACCTTCAACGAGCCCGGCTCTGAGTACTTTCATCTTCTGCTCAGCACCCGG 3546
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2018 CTGTTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAATCATTTATGACAGTGACT 2077
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3547 CTGGGGGCTCGGCTGAACCTCCAGTCGGCAGACACTGTGATCATTTTGGACAGCGACT 3606
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2078 GGAATCCTCATGTGCTGATCTTCAAGCAATGGCTAGAGCTCATCGACTTGGCCAAACAATA 2137
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3607 GGAATCCTCACAGGACCTTGAAGCGCAGGACCGGCCACCGCATCGGGCAGCAACG 3666
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 2138 AGGTGATGATTTATAGGCTCATAAACCGAGGCCACCATTTGAAGAAAGGAT 2186
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 3667 AGGTGCGTGTCTCCGCTCTGCACCGCTCAACAGCGTGGAGGAGAGAT 3715
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: September 19, 2004, 02:55:47

Job time : 274 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 20:33:51 ; Search time 1790 Seconds
(without alignments)
11778.315 Million cell updates/sec

Title: US-10-049-137-1
Perfect score: 4177
Sequence: 1 atgagtagttgttgagag.....caactggtaatacaagattc 4177

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3327077 seqs, 2523723180 residues

Total number of Hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1145.2	27.4	4551	17	US-10-675-072A-42
2	1038.8	24.9	4459	17	US-10-437-963-53427
3	850.8	20.4	3204	9	US-09-938-842A-2526
4	850.8	20.4	3204	11	US-09-938-842A-2526
5	434.2	10.4	2578	13	US-10-424-599-91500
6	421.4	10.1	5739	9	US-09-960-253-142
7	421.4	10.1	6417	9	US-09-962-436-288
8	421.4	10.1	6417	15	US-10-096-534-28
9	421.4	10.1	6475	15	US-10-037-270-325
10	421.4	10.1	6475	16	US-10-117-722-325
11	420.8	10.1	6331	15	US-10-175-523-51
12	420.8	10.1	6331	16	US-10-159-563-176
13	420.6	10.1	6337	16	US-10-062-674-1723
c 14	409.6	9.8	7805	14	US-10-044-090-370

15	409.6	9.8	7805	15	US-10-006-285-400	Sequence 400, App
16	400.4	9.6	854	13	US-10-424-599-23261	Sequence 23261, A
17	387	9.3	3716	16	US-10-104-047-476	Sequence 476, App
18	387	9.3	4675	17	US-10-115-635-216	Sequence 216, App
19	387	9.3	9646	13	US-10-462-261-1	Sequence 1, Appl
c 20	321.2	7.7	7001	17	US-10-437-963-98673	Sequence 98673, A
21	316.2	7.6	612	13	US-10-424-599-111448	Sequence 111448, A
22	315.4	7.6	620	13	US-10-424-599-60603	Sequence 60603, A
23	296.6	7.1	6608	8	US-08-973-363-10	Sequence 10, Appl
24	285.4	6.8	364	14	US-10-071-766-140	Sequence 140, App
25	285.4	6.8	7764	16	US-09-814-353-21548	Sequence 21548, A
26	285.4	6.8	7764	16	US-10-341-434-60	Sequence 60, Appl
27	272.2	6.5	1862	13	US-10-425-114-13362	Sequence 13362, A
28	272.2	6.5	3262	13	US-10-412-698B-1579	Sequence 1579, App
29	269	6.4	1979	13	US-10-424-599-84139	Sequence 84139, A
30	266.2	6.4	4407	16	US-10-369-493-25462	Sequence 25462, A
31	264	6.3	1874	17	US-10-005-057A-1	Sequence 1, Appl
32	264	6.3	1874	17	US-10-675-072A-1	Sequence 5, Appl
33	259.6	6.2	3120	13	US-10-058-024-5	Sequence 5, Appl
c 34	259	6.2	3594	17	US-10-437-963-51978	Sequence 51978, A
35	253.4	6.1	3743	17	US-10-437-963-49246	Sequence 49246, A
36	252.6	6.0	3192	9	US-09-938-842A-2271	Sequence 2271, App
37	252.6	6.0	3192	11	US-09-938-842A-2271	Sequence 115, App
38	249.8	6.0	2195	17	US-10-343-710-115	Sequence 114, App
39	249.8	6.0	6222	16	US-10-120-988-114	Sequence 31425, A
40	248.2	5.9	3548	13	US-10-425-114-31425	Sequence 1583, App
41	248.2	5.9	4011	13	US-10-412-699B-1583	Sequence 25997, A
42	247	5.9	4221	16	US-10-369-493-25997	Sequence 431, App
43	245.4	5.9	3174	10	US-09-934-455-431	Sequence 789, App
44	245.4	5.9	3174	13	US-10-412-699B-789	Sequence 1845, App
45	245.4	5.9	3174	13	US-10-412-699B-1845	

ALIGNMENTS

RESULT 1

US-10-675-072A-42
; Sequence 42, Application US/10675072A
; Publication No. US20040098760A1
; GENERAL INFORMATION:
; APPLICANT: Yumin, Tao
; APPLICANT: Gordon-Kamm, William
; APPLICANT: Shen, Bo
; APPLICANT: Lowe, Keith
; APPLICANT: Danilevskaya, Olga
; APPLICANT: Mahajan, Pramod
; APPLICANT: Rafalski, Jan Antoni
; APPLICANT: Sakai, Hajime
; APPLICANT: Klein, Theodore
; TITLE OF INVENTION: Transcriptional Regulatory Nucleic
; FILE REFERENCES: 12868
; CURRENT APPLICATION NUMBER: US/10/675,072A
; PRIOR FILING DATE: 2003-09-30
; PRIOR APPLICATION NUMBER: 10/005,057
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/251,555
; PRIOR FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 4551
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (343)...(4332)
US-10-675-072A-42

Query Match 27.4%; Score 1145.2; DB 17; Length 4551;
Best Local Similarity 62.1%; Pred. No 3e-295;
Matches 2098; Conservative 0; Mismatches 1193; Indels 87; Gaps 15;

QY 754 CATAAAGAAATCCAGAGACTTTCAACAGTTTGTATCATCTCTGAATTCCTCAAAAGGC 813
D 755 |||||
D 756 |||||
QY 640 AAGAGTGTCACTCGTGAGATACGTCAATATAAGAGAGAGCCGGAAGTTTCTCTCTGGTGGG 699
D 641 |||||
QY 814 TTGTTAATCATACACAGCTTGGAGGACTTAATTTTGGGTTCTCGTGGTCAAAACAG 873
D 815 |||||
D 700 AACTATACCTCACTACAGCTTGAAGGTTAACTTCTTAAGCTTATCTTGGTACCAAT 759
QY 874 ACGCATGTATCTCTGATGAATGGGACTAGGCAAGCAATTCAAAGCATTTGCCCTT 933
D 875 |||||
D 760 AAACGTGTATACCTGGGAGAGATGGTCTTGGTAAACGATACAGAGATTTGGCTTT 819
QY 934 TTAGCTTCACTTTTGGAGGAACTCATTCGGCAATTTGGTAATTTGCTCTCTATCGACT 993
D 935 |||||
D 820 TTGGCTCTCTATTTGGACAGCTTGGTCTCTCATCTGGTCTTGGACCCCTTCAAC 879
QY 994 CTGGTAATCTGGGAGAGAGTTTGGCACATGGGCCCCACAGATGAACGTGGTTATCTAT 1053
D 995 |||||
D 880 TTGGCAATTTGGGAGCGTGAATTTGCAACTTGGGCACCTCAATGAATGTTGTAATGTAT 939
QY 1054 TTTGGCACTGGCAAGCTCGAGAGTTTATCAGAAACATGATTTTACTTATCGAAAGAT 1113
D 1055 |||||
D 940 TTTGGATCTGCAGCTTCTCGTGAATTTATAGGAAGTATGAATTTTACTACCCCAAGAG 999
QY 1114 CAA---AAAAAGATCAAGAAAAAGAAATCTGGACAAATAAGTAGCGAAAGCAAGCAAAAA 1170
D 1115 |||||
D 1000 AAACCTAAGAGCTTAAGAAAAAGAAATCATCTCCATCTAATGAAGCAAGAAACATCA 1059
QY 1171 AGAATCAAGTTTGTATGCTCTCTCATCATGTATGAGATGATCAACCTAGATTTCAAGAGTT 1230
D 1172 |||||
D 1060 AGAATAAAATTTGATGCTCTTTGACATCTTATGAGATGATCAATATGGATTTCTACTGTT 1119
QY 1231 CTAAACCAATTAAGTTGGGAGTGCATGTTTGCATGAAGTCACTGACTGAAATAAG 1290
D 1232 |||||
D 1120 CTAAAAACCATAGAAATGGGAATGATGATTTGATGAGGGGCATCGTTTGAAGAAACAA 1179
QY 1291 GATTCAAAAGCTGTTCTCTTCACTGACACAGTATTTCAAGTAACCAACCGTATTTCTTGACA 1350
D 1292 |||||
D 1180 GATTCGAAGTTGTTGGTCACTTAAGAGATATCATACTAAGCATCGTTCTCTTAACG 1239
QY 1351 GGAACACCACTTCAGAACAACTTGGATGAACCTTTTCATGCTCATGCAATTTTCTTGATGCG 1410
D 1352 |||||
D 1240 GGAACCCAGTTTCAGAACCACTGGATGAACTTTTCATGCTTATGCACTTCTTGGAGGT 1299
QY 1411 GGAAGTTTGAAGTTTGGAGAGTTCCAGAGAGTTTCAAGATATTAATCAAGAGAG 1470
D 1412 |||||
D 1300 GACAGTTTGTGTATAGCTGATCTCCAGAGGAAATTCAGGACATAAAACAGGACAA 1359
QY 1471 CAGATCTCAAGGTTGCACAAATGTTGGCTCCACATTTGCTCAGAAGGGTAAAAAAGAC 1530
D 1472 |||||
D 1360 CAAGTTGAGAAGCTCCATGATGCTGAAGCCGATCTTCTTGGCGATTCAGAAAGAT 1419
QY 1531 GTATGAAGACATGCCCCCAAAAGAGGCTCAATTTGGTGTGTGATCTGAGCAGTCTG 1590
D 1532 |||||
D 1420 GTTATGAAGAACTTCTCCGAAAAAGAGCTGATTTTACGTGTGAGCTTACAAACAA 1479
QY 1591 CAGAAGATATTAACAGCTATTTTACCCGTAATTTCAAGTATTAAGTATTCAGAAAAAGGGA 1650
D 1592 |||||
D 1480 CAAAAGAGTACTATAAGCAATTTCTACCAAGAAATTAAGAGTTTAACTCGCTGAGT 1539
QY 1651 GGTGCTCAAAATTCCTTTTAAACATTTATGATGAATTAAGAAAGTATGCTGCCATCCT 1710
D 1652 |||||
D 1540 GGTGACATGTTTCACTTATTAATGTTGTAATGAACTACAGCAACCTTGTGGCCATGCA 1599
QY 1711 TATATGTAGAGGTGTTGA----- 1730
D 1712 |||||
D 1600 TTCTATGACAGATGAACCGGAGAGCTGCCAAATTCAGAAGAGCTTTTAAGATTTTCTGAT 1659
QY 1731 -----GCCAGTTATTCAGAC----- 1746
D 1660 GTCATTAACATTTTAAATTCACCTACAGCTTTTATTTACTTTCAAAACACATATTGATGCA 1719

QY 1747 ----- 1746
D 1720 GAGGTACTGGAGTTCTCTGATATCTCTGCTGTATCTTTTCACTTCAOGTCTTAACAA 1779
QY 1747 -----CGAATGAAGCTTTCAAAACACTTTTGGAGTCTTTGCGAAAGCTG 1791
D 1780 CACAATGCCATATTTGAAGTTGTGACATCTGGAGGCTTTTAAATCTTCTTGAAAAATG 1839
QY 1792 CAACTTCTAGATAAAATGATGGTCAAACTGAAAGAGCAAGGACACAGAGTCTTAATATAC 1851
D 1840 GAGCTGCTTGACAAAGATGATGGTGAACCTGAAGAGACAGGTCACAGGGTCTTATTTAT 1899
QY 1852 ACACAGTTTTCAGCATATGCTGGACTTACTTGAAGACTACTGTACCATAAGAAATGGCAG 1911
D 1900 TCACAGTTCCAGCACATGTTGGACTTACTTGAAGATTAATTAAGCTACCGGAAATGGAGT 1959
QY 1912 TAGAGCGGAATTCATGGAAGGTTGGCGGAGCTGAGCGCAATATACGATAGATCGGTTT 1971
D 1960 TATGAGCGTATTTGATGGAATAATAGGTGGTGGTCTGAGAGGACAGATACGAATTTGATCGCTTC 2019
QY 1972 AATGCCAAAAATTTCTAACAAGTTTCTTTTCTCTCCACAAAGAGCTGGTGGCTTAGGA 2031
D 2020 AATGCTAAAAATTTCTACTAGTTTCTCTTCTTCAACAGAGCTGGTGGTCTTGGT 2079
QY 2032 ATAAATCTTTCACCGCTGATACAGTAATCAATTTATGACAGTGAATTCCTCATGCT 2091
D 2080 ATAAATCTGCAACTCCGATACCTGTAATTAATTTATGACAGTGAATTCCTCATGCT 2139
QY 2092 GATCTTCAAGCAATGCTAGAGCTCATGACTTGGCCCAACAAATTAAGGTGATGATTTAT 2151
D 2140 GATTTCAAGCTATGCGAGAGCCCATCGCTTAGGGCAGACACAGCAAGGTTATGATATAT 2199
QY 2152 AGCTCATAAACCGAGCGCACTTGAAGAAAGGATGATGCAATTTGACTAAAAAGAAATG 2211
D 2200 AGCTTGTGTAGTCTGCTGTAATGAGGAGCGGATGATGCAACTTACAAAGAAATAATG 2259
QY 2212 GTTCTAGAGCATCTGTTGTTGGAAACTC---AAAAACAAACAACTTAATTCAGGAAGAG 2268
D 2260 GTATTTGGAGCACTTAGTTGTTGGCGCTCTCACGAAAGGCACATAATTTGTCAGAGGAG 2319
QY 2269 TTAGATGACATCATAGGTATGATCAAGAGAGCTTTTGTCTAGTGAAGATGATGAAGCA 2328
D 2320 TTGGATGATATTAATTCGGCATGGCTCAAGGAACTTTTGTACGATGAAATGATGAAGCC 2379
QY 2329 GGAAGCTCTGAAAAATTCATTAATGATGATCGCGCTATAGACAAATTTGCTTGATCGTAT 2388
D 2380 GGAATCTTGCCAAATCCATTAATGATGATGCTGCGATTTGATAGATTTATGGACCGTAC 2439
QY 2389 CTCGTGAGGACAGAGGAAGTCTCAGTGGATGATGAAGAGAGAAATGGAATTTTAAAGGCT 2448
D 2440 CAAAGCTGA---CGGGCAAGAGCCTGTGGAAGATGAAGAGACAGAAATTTCTAAAAAGGA 2496
QY 2449 TTCAAGGTGGCTAAATTTTGAATATATAGATGAATAAGTGAAGGACAGCATTTAGAGGCACAG 2508
D 2497 TTCAAGGTGCTAACTTGAAGTACATAGATGAAGCTTAAAGCTTTAGCTGCAAAAGAGGAG 2556
QY 2509 AGAGTCTGCTCTCAAAAGCAAACTCTCAGCAGGCAATTTCTGATAGAGCAAGTATTTGGAA 2568
D 2557 GAGCGACGCAAAAGGCCGAGCTGAAGCTGCAAACTCTGACAGAGCAAAATTTTGGGAT 2616
QY 2569 GAGTTGTTTAAAGATAAATTTGAGCTGCAAGGCTGAGAGCTTAAATGCTTCTTGGAAAA 2628
D 2617 AAATTTTGAAGGACAGATACGATGTACAAAAAGTTGAAGAGCATACTACTATGGGAAAA 2676
QY 2629 AGAAGAGAGTCCGACAGCTGGTATCCATTTGAAGAGATGATCTTGTGCTTTGGAA 2688
D 2677 GGAAGAGGAGCGCCAAACAGATGGTGGTCTGATGAGGACGACATTTACCGGCTGCTAT 2736
QY 2689 GATGTGAGCTCT---GATGGAGATGAAGTTTATGAAGCTGAGTCAACAGATGTTGAAGCA 2745
D 2737 GACATGAGTCTGAAGATGATGATTTACTCTTATGATGATGATGCTGTCAGATACGACACA 2796
QY 2746 GCAGGCAAGGA-----GTTTCAGCGGTCGACGGCGGTACAGAGAAAGGGTTCGGAT 2799

QY 1405 GATCGGGGAAGTTTGGAGTTTGGAGAGTTCCAGGAGGAGTTTCAAGATATTAATCAA 1464
 Db 985 ----- 984
 QY 1465 GAGGAGCAGATCTCAAGGTTTGCAAAATGTTGCTCCACATTTGCTCAGAAAGGTTAAAA 1524
 Db 985 -----AAA 987
 QY 1525 AAAGACGTATGAAGACATGCCCCCAAAAGAGAGCTCAATTTGCGTGTGTGATCTGAGC 1584
 Db 988 GACGTGCTCAAGGATAAGGTGCCCCCAAGAGAGAGCTCAATTTGCGTGTGTGATGATGAGC 1047
 QY 1585 AGTCTGAGAGAAATATTAAGAGCTATTTTACCGGTAAATATCAAGTATTAAGTATGACAAA 1644
 Db 1048 AGCCAGCAAAAGAGGTGTAAGAGCTGTATTAACCAATATTAACAGTCTTGACAAAG 1107
 QY 1645 AAGGAGGTGCTCAAAATTTCCCTTAATAACATATATGATGAAATTTACGAAAAGTATGCTGC 1704
 Db 1108 AACGAGATGCTAAATATC-----TAACTGCTATGAATTTACGCCAAGTTTGTCT 1161
 QY 1705 CATCTTATATGCTAGAGGTTGAGCCAGTATTAACAGACGCAATGAAGCTTTCAAA 1764
 Db 1162 CATCCGTATCTGTACCAGACTTTGAGCCAAAGATTTGAAGACGCAAAATGAAGCTTTTACA 1221
 QY 1765 CAACCTTTGGAGTCTTGGAAGCTGCACTTCTAGATAAAATGATGTCMAACTGAAA 1824
 Db 1222 AAATATTTGGAAGCTTCGGAAGGTGCACTGTTAGCAAAATGATGGTCAAACTCAA 1281
 QY 1825 GAGCAAGACACAGAGCTCTAATATACACAGATTTTCAGCATATGCTGGACTTACTTGAA 1884
 Db 1282 GAGCAGGGCACAGAGTCTAATCTATACCAATTTCCACACACCTTTTACCTTTTGGAA 1341
 QY 1885 GACTACTGTACCCATAGAATGCGAGTACAGCGAATTTGATGAAAGGTTGGCGAGCT 1944
 Db 1342 GACTACTTCTCAAGATTTGAATTTATGAACGATTTGATGAAAGATCAGTGGACCT 1401
 QY 1945 GAGCGCAATACCATAGATCGTTCATGCGCAAAATTTCTACAAATTTTGTGTTTTG 2004
 Db 1402 GAGAGCAAGTAAGAATAGATCGGTTTAATGCAAAATCTGAAACCGGTTCTGTTTTG 1461
 QY 2005 CTCTCCACAGAGCTGGTGGTTAGGAATAAATCTTCAACGGCTGATACAGTAATCAT 2064
 Db 1462 CTGATCAAGAGCTGGTGGTAAGGATAAATCTTCAACAGCTGATACAGTATCAT 1521
 QY 2065 TATGACAGTACTGGAATCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATCGACT 2124
 Db 1522 TATGACAGTATTTGGAATCTCATGCTGATCTCCAGGCTATGGCTAGAGTTTCACTCGGCT 1581
 QY 2125 GGCACAAACAAATAGGTGATGATTTATAGGCTCATAAACGAGGCACCATTTGAAGAAAG 2184
 Db 1582 GGCACAAACAAATAGGTGATGATTTATAGGCTCATAAACGAGTCTGATAGAAAGACGG 1641
 QY 2185 ATGATGCAATTTGACTAAAAAGAAATGGTCTAGAGCATCTTGTGTTGGGAAACTCAA 2244
 Db 1642 ATGATGGAGATAACCAAGAACAGATGCTTCTAGAGCATCTCGTCTGGG-----A 1692
 QY 2245 ACACAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATGATCAAGAGGACT 2304
 Db 1693 AAACAAACATCTTTGTCAGGACGAGTTAGATGACATCATCAATATGTTTCCAGAGAACT 1752
 QY 2305 TTTCTAGTGAAGATGATGAGCAGGAAGTCTCGGAAAAATTTCAATATGATGATGCGGCT 2364
 Db 1753 TTCTCTGAAGAGATGATGAAGCAGGAGGTTCTGAAAAATTCATATGATGATGCTGCT 1812
 QY 2365 ATAGACAAATTTGCTTGTATGCTGATCTCGTGAGGACAGGAAGTCTCAGTGGATGATGAA 2424
 Db 1813 ATCGACAGTGTCTAGATCGTAATCATGATAGTCTGTGAGGTTCTCACTGGATGATGAG 1872
 QY 2425 GAGGAGATGATTTTAAAGGCTTTCAAGGTGCTAATTTTGAATATATAGATGAAAT 2484
 Db 1873 GAGGAAACTGATTTTTTAAAGAAATTTCAAGGTGCTAGTTTGAATGATGATGATGAA 1932

QY 2485 GAGGACAGCATTAGAGGCACAGAGTCGCTGTGAAAGCAAAATCTTTCAGCAGGCAAT 2544
 Db 1933 AATGAGCTGCAGCATTTAGAGGGCCAAAGCTATAGAAAACAATTTCTTTCAGTAAGAAAT 1992
 QY 2545 TCTGATAGAGCAAGTTTATTTGGGAAGAGTTGTTTAAAGATAAAATTTGAGCTGCAACGAGT 2604
 Db 1993 GCTGATAGAACAAAGTCACTGGAAAGATTTGCTGAAAGACAAATATGAAGTGCACAAAGCT 2052
 QY 2605 GAGGACCTTAATGCTCTTGGAAAAAGGAAGAGTCCGAAGCAGTTGTTATCCATTGAA 2664
 Db 2053 GAGGAGCTCAGTCTCTTGGAAAGAGGAAGAACCGCAAGCAGGTGATGTA---TGCT 2109
 QY 2665 GAAGATGATCTTCTGTTTGGAAAGATGTGAGCTCTGATGGAGATCAAAAGTTATGAAGCT 2724
 Db 2110 GAAGATGATCTTCTGTTTGGAAAGATGTGAGCTCTGATGGAGGACCAATATTTCTTGT 2169
 QY 2725 GAGTCAACA-----GATGGTGAAGCAGCAGGACAA 2754
 Db 2170 GATTTAAAGTGAACCTCTGATGAGGAGGAGAGCAGATGAACCCGAAGCAGCTCGACAG 2229
 QY 2755 GAGTTTC---AGACGGGTGACGGCCGTACAGAAAGAGGGTCCGATATTTTGGAAACCA 2811
 Db 2230 AGAAAGCCGAGGACGTTTCAAGGCCCTTACAGAAAAAGGGCTCGCGATATTTTCGGAGGAA 2289
 QY 2812 ACTCCGTTGATGGAAGGTGAGGGAGATCTTTTCAAGAGTACTGGGTTTCAACAGAGTCAA 2871
 Db 2290 ATACCTTTTATGGAAGCGAGGAGATATCTATGCTGCTGGATTTTAAACGAGACTGAG 2349
 QY 2872 AGGGCCATTTTGTACAGACTTTTATGAGGTATGAGAGTGGCAATTTTGAATGGAAGGAG 2931
 Db 2350 AGGGATATATTTTACGTACGTTTAAAGAGGTATGGAATCTTATTTTGAACACATTTGCC 2409
 QY 2932 TTTGTTCTCGTTTAAAGCAGAGACCTTTTGAAGAAATAAATAATGAAATGGAATACTCTTC 2991
 Db 2410 GAAACCCCTACCGATAATTTTCAAAATTTTAAAGTTTATACGGCAATGGT----- 2459
 QY 2992 TTGAAGCACATTTGCTGAAGAAATAGACGAGAAATTTCTCAACCTTTTTCAGATGGTGTGCC 3051
 Db 2460 -----CTATGACAGATGAGTTCCC 2478
 QY 3052 AAGGAAGGACTTTAGAATAGAAGATGTTTAGTCAGAAATTTGCTTCTCTGATATCTAGTTTCA 3111
 Db 2479 AAAGAAGGATAGACAGTACGAGCTACTAGTGAGCATGACTTTTATGATGCTAGTAAAG 2538
 QY 3112 GAGAAGTGAAATTTCTAGAAATCATCCAGGGAACCTGTTTCCCTCTCTGCAATCTT 3171
 Db 2539 GAGAAGTGCAATTTTGGACAAACCATCCGACAGCACCTGTTTTT---CTCTAACTAGTC 2595
 QY 3172 GAAAGATTTCCCGGACTGAGAAAGTGGAAAAATTTTGAAGGAGGAAACATGACAGATAATG 3231
 Db 2596 ATCAGTAAGTACAAATTTGAGAAATGAGCATTTTCTAAGGAAGAAACATGATAGATACIG 2655
 QY 3232 ATACGTGCTGTTTAAAGCATGGGTACGAGCGGTGGCAAGCTTATTTGATGACAAAGAG 3291
 Db 2656 ATTCTGCTGTTTCCAAAGCATGCTATGTTAGTGGTGGCCATCGTTTGAAGACGAAGAG 2715
 QY 3292 TTGGGATCCAGAGCTTATCTGCAAGAAATGAAATTTCCCT 3333
 Db 2716 ATAGGGTTCGAAGAGTTGCTTGCCTGCAAAAGACCTGAAATATCCCT 2757

RESULT 4
 US-09-938-842A-2526
 ; Sequence 2526, Application US/09938842A
 ; Publication No. US20040009476A9
 ; GENERAL INFORMATION:
 ; APPLICANT: Harper, Jeff
 ; APPLICANT: Krepes, Joel
 ; APPLICANT: Wang, Xun
 ; APPLICANT: Zhu, Tong
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 ; TITLE OF INVENTION: SAME, AND METHODS OF USE
 ; FILE REFERENCE: SCRIPI300-3

; CURRENT APPLICATION NUMBER: US/09/938,842A		
; CURRENT FILING DATE: 2001-08-24		
; PRIOR APPLICATION NUMBER: US 60/227,866		
; PRIOR FILING DATE: 2000-08-24		
; PRIOR APPLICATION NUMBER: US 60/264,647		
; PRIOR FILING DATE: 2001-01-16		
; PRIOR APPLICATION NUMBER: US 60/300,111		
; PRIOR FILING DATE: 2001-06-22		
; NUMBER OF SEQ ID NOS: 5379		
; SEQ ID NO 2526		
; LENGTH: 3204		
; TYPE: DNA		
; ORGANISM: Arabidopsis thaliana		
; US-09-938-842A-2526		
Query Match 20.4%; Score 850.8; DB 11; Length 3204;		
Best Local Similarity 61.0%; Pred. No. 1.9e-216;		
Matches 1892; Conservative 0; Mismatches 697; Indels 513; Gaps 11;		
QY	280 GTTAGTCTCTTAACGAGATAGATAGATATGGATGTGGAATGCGTCTCAAAATCT	339
DB		
QY	121 GATTGCTCTTGGCGAGATTGAAAGATTGGACCGTGAATGGCGTCTACTGCATCT	180
DB		
QY	340 AGTGAAACAAGTCTCCCGAT---GCGGAACCGAAGCCAAATTTTGTGAAACAGATCTC	396
DB		
QY	181 ACAATCCGAATCTCTGACAAACGGGACCAACCCCTAGTTGCGTGAACAATATCTT	240
DB		
QY	397 GTGAAGTGAAGGATTATCATACCTTCACTGCTCTTTGGGTGCTCGAAGGAGTTCCAG	456
DB		
QY	241 GTGAGTGAAGGGTTATCGTACTTGCATGCTCTTTGGTCCCGAGCAGGATTCGAG	300
DB		
QY	457 AAGCTTTAAGTCAATCATCGTT-----TAAAAACCAAGATGAACAAATTTTCAACCGT	510
DB		
QY	301 AAGCTTTACAAGTCCCATCTCTATTAAAACTAAAAATTAAGGGTGACCAAGTTTAAACGCT	360
DB		
QY	511 CAATGGAGTCAATCA-----ATACAGCGAAGATGATTTTGTGCCATACGTCCTGAG	564
DB		
QY	361 GCTATGGACGATTTATCGCTGAGAACGGAGCTCATGAGTTTATCGCCATTCGTCCTGAA	420
DB		
QY	565 TGACCACTCTTGATCGGATTTCTGCTGCAGAGAGGAAGATGGGAGCTTGGAAATATCTT	624
DB		
QY	421 TGGAAACTGTTGACCGGATTTATGCTTGGCGGAGGAGATGATGGGAGGAGTATCTG	480
DB		
QY	625 GTCAATATAAAGAGCTATCCTATGATGAATTTATGGAGATCAGAAATCAGACATCTCA	684
DB		
QY	481 GTGAAGTACAAGAACTTTCTTACAGAAATAGTTTACTGGGAACTCTGAATCTGACATCTCA	540
DB		
QY	685 ACCTTCCAGATGAATTCAGAGTTCAAGGATGTAAATCTAGAACTCGCAGAACTAAA	744
DB		
QY	541 GATTTCCAGAAATGAAATCCAAAGGTTTAAAGGATATAAACTCTAGCTCTCGTAGAGATAA	600
DB		
QY	745 GATGTTGACCATAAAGAAATCCAGAGACTTTCAACAGTTTGATCATATCTCTGAAATTC	804
DB		
QY	601 TATGTCGAAATGAGAAATCAGAGGAATTCAAACAGTTTGATCTTACTCTCGAATTT	660
DB		
QY	805 CTCAAGGCTGTTTACATCATACAGCTTTAGGGACTTAATTTTTCGGGTTCTCGTG	864
DB		
QY	661 CTTACAGGCATTTGCATACATACAGCTCCAGGCTGAAATTCCTAAGGTATTCATGG	720
DB		
QY	865 TCAAAACAGACGATGTAATCTCTGCTGATGAATGGGACTAGGCAAGCAATTCAAAGC	924
DB		
QY	721 TCCAAAGAAACCAATGTAATCTCTTGGCGATGAAATGGGACTAGGCAAACTATTTCAGAGC	780
DB		
QY	925 ATTGCCCTTTTAGCTTCACTTTTGTAGGAGAACTCATTCCTCGCATTTGGTAAATTCCTCT	984
DB		
QY	781 ATTGCCCTTTTGGCTCTTTTGTTCGAGGAAACTTATCTCTCATCTAGTTTGTCTCT	840
DB		
QY	985 CTATCGACTCTGCGTAACTGGGAGAGAGTTTGCACATGGGCCCCACAGATGAACGTG	1044
DB		
QY	841 CTCTCAACCATCCGTAACTGGGAGAGGAAATTTGCCACCTGGGCACACATATGAATGTG	900
DB		
QY	1045 GTTATGTATTTTGGCACTCGGCAAGCTCGAGCAGTTTATCAGAGAAATGATGTTTACTTA	1104
DB		

DB	901 GTTATGTACACTGGGGATTTCGGAAGCAGCTGATGTTATATATGGAAACATCAGTTCTACTTC	960
QY	1105 TCGAAAGATCAAAAAAGATCAAGAAAAAGAAATCTGGCAAAATAGTAGCGAAAGCAAG	1164
DB		
QY	961 TCTGAAGGGCGTAAAAAGGTTAAAG-----	984
DB		
QY	1165 CAAAAAGAAATCAAGTTTGATGTCTCTCCATCATCGTATGAGATGATCAACCTAGATTCA	1224
DB		
QY	985 GCAGTTCTAAAAACCAATTAAGTGGGAGTGCATGATTGTTGATGAAGGTTCATCGACTGAA	1284
DB		
QY	1285 AATAAGGATTCAAAAGCTGTTCTCTTCAATTGACACAGATATTCAAGTAACCAACCGTATTCTT	1344
DB		
QY	1345 CTGACAGGAAACACCACCTTCAGAAACAACTTGGATGAACCTTTTTCATGCTCATGATTTCTTT	1404
DB		
QY	1405 GATCGGGGAAGTTTGGAAAGTTTGGAGGAGTTCCAGAGGAGTTCAAAGATATTAATCAA	1464
DB		
QY	1465 GAGGAGCAGATCTCAAGGTTGCACAAATGTTGGCTCCAATTTGCTCAGAGGGTAAAA	1524
DB		
QY	1525 AAAGAGTAAATGAAGACATGCCCCCAAAAAAGGAGCTCAATTTTGGTGTGATCTGAGC	1584
DB		
QY	988 GACGTGCTCAAGGATTAAGGTGCCCCCAAAAGAGGAGCTCAATTTTGGTGTGATGATGAGC	1047
DB		
QY	1585 AGTCTGCAGAAAGATATTACAAAGCTATTTTACCCTGTAATTTATCAAGTATTGACAAA	1644
DB		
QY	1048 ACCAGCAAAAAGAAAGTGTAACAAAGCTGTTATCAATTAATTAACAGTCTTGACAAAG	1107
DB		
QY	1645 AAGGAGGTGCTCAAAATTTCCCTTAATTAACATTTATGATGAATTAAGAAAGTATGCTGC	1704
DB		
QY	1108 AAACGAGATGCTAAATATC-----TAACGTGCTAAATGAATTAACGCAAGTTTCTCT	1161
DB		
QY	1705 CATCTTATATGCTAGAGGGTGTGAGCCAGTTTATCAGCGCAAAATGAAGCTTTCAAA	1764
DB		
QY	1162 CATCGGTATCTGTTACCAAGCTTTGAGCCAAAGTTTGAAGAGCGCAATGAAGCTTTTACA	1221
DB		
QY	1765 CAACCTTTTGGAGTCTGTGGAAGCTGCAACTCTTAGATAAAATGATGTCAAACTGAAA	1824
DB		
QY	1222 AAACCTTAATGGAAGCTTCCGAAAGTTGCAACTGTTAGACAAATGATGTCAAACTCAA	1281
DB		
QY	1825 GAGCAAGGACACAGAGTCTCTAATATACACACAGTTTTCAGCATATGCTGGACTTACTTGAA	1884
DB		
QY	1282 GAGCAGGGGACACAGAGTGTCTAATCTATACCAATTCACACACACCTTTTACCTTTTGAA	1341
DB		
QY	1885 GACTACTGTACCCATGAAGAAATGGCAGTAGACAGCAATTTGATGGAAGGTTGGCGAGCT	1944
DB		
QY	1342 GACTACTTCACTTTCAAGAAATTTGAAATTTATGAACCGATTTGATGGAAGATCAGTGGACCT	1401
DB		
QY	1945 GAGCGCAAAATACGATAGATCGTTCAATGCCCCAAAAATTTCTAAACAGTTTGTGTTTTTG	2004
DB		
QY	1402 GAGGCAAGTAGTAAGTAAGTAGATCGTTTAAATGAGAAACTCGAACCGTTTCTGTTTTTG	1461
DB		
QY	2005 CTCTCCACAAGAGCTGGTGGCTTAGGAATTAATCTTGAACGGCTGATACAGTAATCATTT	2064
DB		
QY	1462 CTCAGTACAAGAGCTGGTGGATAGGATAAACCTTGCAACAGCTGATACAGTCATCATTT	1521
DB		
QY	2065 TATGACAGTGAATGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATGAGCTT	2124
DB		
QY	1522 TATGACAGTGAATGGAATCCTCATGCTGATCTCCAGGCTATGGCTAGAGTTTATCGGCTT	1581
DB		
QY	2125 GGGCAAAACAAATAGGTGATGATTTATAGGCTCATAAACCGAGGCCACCTTCAAGAAAGG	2184
DB		
QY	1582 GGGCAAAACAAATAGGTGATGATCTACAGCTGATTTCAAAAGGTACTGTGAAGAAACGG	1641
DB		

```
QY 2185 ATGATGCAATTGACTAAAAAGAAATGGTTCTAGAGCATCTTGTTGTTGGGAAACTCAA 2244
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1642 ATGATGAGATAACCAAGAACAGATGCTCTAGAGCATCTGTCGTGGG-----A 1692
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2245 ACACAAACATTAATACGGAAGAGTTAGATGACATCATCAGGTATGGATCAAAAGGAGCTT 2304
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1693 AAACAACATCTTTGTCAGGACGAGTTAGATGACATCATCAAAATATGTTCCNAGGAACCT 1752
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2305 TTTGCTAGTAGAGATGAAGCAGGAAGTCTGGAAAAATTCATTATGATGATGCGGCT 2364
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1753 TTCTCTGAAGAGAATGATGAAGCAGGAGGTCTGAAAAATCCATTACGATGATGCTGCT 1812
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2365 ATAGACAAATGCTTTGATCGTGATCTCGTGAGCAGAGAGTCTCAGTGGATGATGAA 2424
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1813 ATCGACAGTTGCTAGATCGTAATCATGTAGATGCTGTGAGGTTCTCCTGGAATGATGAG 1872
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2425 GAGGAGAAATGATTTCTTAAAGGCTTTCAAGGTGCTAAATTTTGAATATATAGATGAAAT 2484
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1873 GAGGAACATGATTTTAAAGAAATTCAGGTGCTAGTTTGTAGTATGTAGATGATGAA 1932
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2485 GAGCAGCAGCATTAGAGGCACAGAGAGTCTGCTGAAAGCAAAATCTTCAGCAGGCAAT 2544
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1933 AATGAGCTCGAGCAATTAGAGAGGSCACAAGCTATAGAAAAACRAATCTTCAGTAAGAAAT 1992
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2545 TCTGATAGACAAGTTATTGGGAAGAGTTGTTAAAAAGATAAAATTTGAGCTGCACAGGCT 2604
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1993 GCTGATAGACAAGTCACTGGAAGATTTGCTGGAAGACAATATGAAGTGCAACAAGCT 2052
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2605 GAGGAGCTTAATGCTCTTTGAAAAAGGAAGAGTCGCAAGCAGTTGGTATCAATTGAA 2664
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2053 GAGGAGCTCAGTCTCTTGGAAAGAGGAAGAAACGCAAGCAGGTGATGTA---TGCT 2109
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2665 GAAATGATCTTGCTGGTTGGAAAGATGTGAGTCTCTGATGAGAGATGAAGTTATGAAGCT 2724
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2110 GAAGATGATCTTGATGGTTTAGAAGAAATCTCTGATGAGGAGGACGAATATTTGCTTGAT 2169
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2725 GAGTCAACA-----GATGCTGAAGCAGCAGGACAA 2754
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2170 GATTTAAAGTGACCTCTGATGAGAGAGGAAGCAGATGAACCGAAGCAGCTCGACAG 2229
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2755 GGAGTTC---AGACGGGTTCGACGGCGGTACAGAGAAAGGGTTCGCGATAATTTTGGAAACCA 2811
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2230 AGAAAGCGAGGACGGTTACAAGSCCTTACAGAAAAAGGGCTCGCGATAATTCGGAGGAA 2289
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2812 ACTCGTTGATGGAAGGTGAGGAGATCTTTACAGATACGTGGTTTCAACCAGAGTCAA 2871
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2290 ATACCTTTGATGGAAGCGAAGGAGATATCTCATGTGTGCTGGGATTTAAACGAGACTGAG 2349
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2872 AGGCCCATTTTGTACAGACTTTGATGAGGTATGGAGCTGSCAAATTTTGTATGGAAGGAG 2931
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2350 AGGATATATTTTACGTACGTTTAAAGGTATGGAATCTTATTTTGAACACATTTGCC 2409
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2932 TTTGTTCTCTGCTTAAAGCAGAGACCTTTTGAAGAAATAAATGAATATGGAATACTCTTC 2991
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2410 GAAACCCCTACCGATAATCTTACAAATTTTAAAGTTATTACGGCAATGGT----- 2459
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2992 TTGAAGCACATTCGTGAAGAAATAGACAGAAATTTCTCAACCTTTTCAGATGGTGTGCC 3051
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2460 -----CTATGCAGATGGAGTTCCC 2478
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3052 AAGGAAGGACTTAGAATAGAGATGTTCTAGTCAGAAATTCGCTTCTGATACTAGTTTCAG 3111
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2479 AAGAAGGGAATAACAGCTGACGAGCTACTAGTGAGCATGACTTTTATGATGCTAGTAAAG 2538
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3112 GAGAAGGTGAATTTGTAGAAAGATCATCCAGGGGAAACCTGTTTTCCCTCTCGCATTTCTT 3171
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2539 GAGAAGTGTCAATTTTGGACAACCATCCGACAGCACTGTTTT--CTCTAACTACGTC 2595
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3172 GAAAGATTCCTCCGACTGAGAAGTTGGAATAAATTTTGAAGGAGGAACATGACAAAGTAATG 3231
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2596 ATCAGTAAGTACAATTTGAGAAATGGAGCATTTTCTAAGGAAGAACAATGATAGATACTG 2655
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 3232 ATACGTGCTGCTTTTAAAGCATGGTACGACGGTGGCAAGCTATTGTTGATGACAAAGAG 3291
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2656 ATTCTGCTGCTTTTCCAAAGCATGGCTATGTTAGATGGTGGCCATCGTTGAAGACGAAGAG 2715
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3292 TTGGGATCCAAAGAGCTTATCTGCAAAAGAAATTCGAATTTCCCT 3333
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2716 ATAGGGTTCCAAAGAGTTGCTTCCAAAGACCTGGAATATCCCT 2757
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
US-10-424-599-91500
; Sequence 91500, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 91500
; LENGTH: 2578
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(2578)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53638C.1
US-10-424-599-91500

Query Match 10.4%; Score 434.2; DB 13; Length 2578;
Best Local Similarity 64.0%; Pred. No. 7.4e-105;
Matches 729; Conservative 0; Mismatches 388; Indels 22; Gaps 4;

QY 2258 ATCAGGAAGATTAGATGATCATCATCAGGTATGGATCAAAAGGAGCTTTTGTAGTGAAG 2317
DB 20 AACAGGAAGATTGGATGATCATCAAGATATGGCTCAAAGGAGTTATTTTCAGATGAGA 79
QY 2318 ATCATGAGCAGCAAAAGTCTGGAATAATTCATTATGATGCGCTATAGACAAATTGC 2377
DB 80 ATGATGAAGCAGCAAAATCCGCCAAATTCATTATGATGCTGCTATTTAGATGATGC 139
QY 2378 TTGATCGTATCTCTGTTGGAGCAGAGGAAGTCTCAGTGGATGATGAAGGAGGAATGGAT 2437
DB 140 TGGATCGTATCAAGTTGGAGATGAAGAGGCTACTTTGGATGATGAAGATGAAGTGGAT 199
QY 2438 TCTTAAAGGCTTTCAAGGTGGCTTAATTTTGAATATATAGATGAAATAGAGGAGCAGCAT 2497
DB 200 TTCTGAAGGCTTTTAAAGGTTGCAAAATTTTGAATATGTTGATGAGGCTGAGGCTGCAAG 259
QY 2498 TAGAGCAGCAGAGAGTCTGCTGTAAGCAAAATCTTCAGCAGGCAATCTCTGATAGACAA 2557
DB 260 AGGAGGAGCAGCAAAAAGAGC-----AATGGAGACACTTAAACAGTTTCAGAAAGGACAC 313
QY 2558 GTTATTGGGAAGAGTTGTTAAAAAGATAAAATTTGAGCTGCAACCGGCTGAGGAGCTTAATG 2617
DB 314 ATTACTGGGAGGATTTGTTTAAAGAGACAGATATCAAGAGCATAAAGTTGAGGAGTTTAATG 373
QY 2618 CTCTTGAATAAAGAGAGAGAGTTCGCAAGCAGTTGGTATCCATTTGAAGAAAGATGATCTTG 2677
DB 374 CCCTGGGCAAGGGGAAACCGGAAAGTTGATGTTTCCGTTGGAGGAGGATGACCTTG 433
QY 2678 CTGGTTTGAAGATGTGAGCTCTGATGG-----AGATGAAAGCTTATGAAGCTGAGTCAACAG 2734
DB 434 CTGGTCTGGAAGATGTAAAGCTCTGATGGTGAAGATGACAAATTTATGAAGAGAGGCTTACTG 493
QY 2735 ATGTTGAAGCAGCAGGACAGAGGAGTTTCAGACGGGTTCAGACGGCGGTACAGAAAGAAA----- 2789
```


Db 494 ATGCTGATTCAAATTCACCTGGAACCTGCTAGAGACCTTATATAAACAAGAGCTC 553
Qy 2790 -----GGGTCCGCGATAAATTTGGAAACCAACTCCGTTGATGGAAGGTGAGGGAGATCTTTC 2844
Db 554 TTACCGCTCGGATAGCACAGAGCACTTCTCTGATGGAAGGTGAAGGAAAGCAATTC 613
Qy 2845 AGAGTACTGGGTTTCAACCAAGAGTCAAAAGGCCAATTTTGTACAGACTTTGATGAGGTAT 2904
Db 614 AGAGTCTTGGTTTAAATCAAAATCAGAGGCTGCATTTGTGCAAAATTTTGTGAGGTTT 673
Qy 2905 GGAGCTGGCAATTTTGAATGGAAGGAGTTTGTCTCCGCTTAAAGCAGAAAGACTTTGAA 2964
Db 674 GGGTGGTGATTTGATTTGAAGGAGTTTACTTCCCGATGAAACAGAAAGACTTATGAA 733
Qy 2965 GAAATAAATGAATGGAATACTCTTCTTGAAGCACATTTGCTGAAGAAATAGACGAGAA 3024
Db 734 GAAATCAAGACTATGAAACCCCTTTCTGTCTCATATTGCTGAAGATATAAATCTG---AT 790
Qy 3025 TCTCCAACCTTTTCAGATGTTGTCGCCAAGAGGACTTGAATAGAAGATGTTCTAGTC 3084
Db 791 TCCGCTACATTCAGATGTTGTTCCAAAGAGGACTCCGAATCCAAAGATGACTTGT 850
Qy 3085 AGAATTCCTCTCTGACTAGTTTCAGGAGAGTGAAATTTGTAGAGATCATCCAGGG 3144
Db 851 AGAATTCAGTCTCTCTTGATAAGGACAAAGTGAAATTTGTGTACAAATCTCTCAA 910
Qy 3145 AAACCTGTTTCCCTCTCGCATTTTGAAGATTTCCCGGACTGGAAGTGGAATAAT 3204
Db 911 ACTCCATAATTTTCAGATGATATATTATACGAAGCGAGTATCTTGCCCTGCCATGAT 970
Qy 3205 TGAAGAGGAGAACATGACAAGATGATGATGCTGCTGTTTAAAGCATGGGTACGGACGG 3264
Db 971 TGAAGAGCGTGCCTCCCTAATGATGCTCCACTTCGCGCATGTACAATCAATAGCTTCAG 1030
Qy 3265 TGGCAAGCTATTGTTGATGACAAGAGTTGGGATCCAGAGCTTATCTGCAAGATTTG 3324
Db 1031 TTGAATTTTCACTGGAATTCGAATGGGATGAAGATCAGTAGGTTCATATCCCTAGTTG 1090
Qy 3325 AATTTCCCTCACATAAGTTTGTCTGCTGCTGAACAAGCTGTTTGAAGGGGCAAGATGG 3383
Db 1091 AATCTTCTTAATAACTGACCAATACAGATGAATTAATTTCCAGGCTCATTTTGG 1149

RESULT 6

US-09-960-253-142
; Sequence 142, Application US/09960253
; Patent No. US20020123619A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.556
; CURRENT APPLICATION NUMBER: US/09/960,253
; CURRENT FILING DATE: 2001-09-20
; NUMBER OF SEQ ID NOS: 187
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 142
; LENGTH: 5739
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-960-253-142

Query Match 10.1%; Score 421.4; DB 9; Length 5739;
Best Local Similarity 51.9%; Pred. No. 3.4e-101;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;
Qy 818 TACATCCATACCAGCTTGAGGACTTAATTTTGGCGTTCTCGTGGTCAAAACAGACGC 877
Db 2177 TGCACCCCTATCAAAATGGAGGGCTGAATTTGGTTGCGCTTCTCTGGGCTCAGGGCACTG 2236

Qy 878 ATGTAATCCTTCTGATGAAATGGGACTAGGCAAGACAATTCAAAGCATTTGCCCTTTAG 937
Db 2237 ACACCATCTTGGCTGATGAGATGGGCTTTGGGAAAACTGTACAGACAGCAGTCTTCTGT 2296
Qy 938 CTTTCACTTTTGAAGGAG-----AACCTCATTCGCCGATTTGGTAATTTGCTCTCTATCGA 991
Db 2297 ATTTCCTTTTACAAGAGGGTTCATTCCAAAGGCCCTTCTCTAGTAGGGCCCTCTTTCTA 2356
Qy 992 CTCTCGGTAACTGGGAGAGAGATTTGGCCACATGGGCCCCACAGATGAAGCTGGTTATGT 1051
Db 2357 CCATCATCACTGGGAGCGGAGTTTGAATGTGGGCTCCAGACATGTATGTGTACCT 2416
Qy 1052 ATTTTGGCACTGGCAGCTCGAGCAGTTATCAGAGAAATGATGATTTTACTTATCGAAG 1111
Db 2417 ATGTGGGTGACAAGGACAGCCGTGCCATCATCCGAGAGATGATGTTCTCTTTTGAAGACA 2476
Qy 1112 ATCAAAAAAGATCAAGAAAAAGAAATCTGGACAAATTAAGTAGCAAGAAAGCAAAAAA 1171
Db 2477 ATGCCATTCTGGTGGCAAGAGCCCTC-----CCGCATGAAGAAAGAGGCAT 2524
Qy 1172 GAATCAAGTTTGTCTCTCTCACATCGTATGAGATGATCAACCTAGATTCAGCAGTTTC 1231
Db 2525 CTGTGAAATTCATGCTGTGCTGACATCTTATGAATGATCACCATTGACATGGCTATTT 2584
Qy 1232 TAAAAACAATTAAAGTGGGAGTGCATGTTGATGAAGTGCATCGACTGAAAAAATAGG 1291
Db 2585 TGGGCTCTATTGATGGGCTGCTCATCGTGGATGAAGCCCATCGGCTGAAGAACAATC 2644
Qy 1292 ATTCAAAGCTGTTCTCTTCAATGACACAGTATTCAAAGTAAACCCGTTATCTTCTGACAG 1351
Db 2645 AGTCTAAGTTCTCCGGTATTTGAATGGTTACTCTCCAGCAACAGCTGTTGCTGACTG 2704
Qy 1352 GAACACCACTTCAGAAACAATTTGGATGAACTTTTTCATGCTCATGCAATTTTCTGATCGG 1411
Db 2705 GGACACCAATTCAAAAACAATCTGGAAGAGTTGTTTTCATCTGCTCAACTTTCTCACCCCG 2764
Qy 1412 GGAAGTTTGAAGTTTGGAGGAGTTCAGGAGAGTTCCAGAGAGTTTCAAGATATTAATCAAGAGAGC 1471
Db 2765 AGAGGTTTCCACATTTTGAAGGTTTTTTGGAGGAGTTTCTGCTGACATTTGCCAAGAGGACC 2824
Qy 1472 AGATCTCAAGGTTGCAAAAAATGTTGGCTCCACATTTGCTCAGAAGGTAAAAAAGACG 1531
Db 2825 AGATAAAAAAATGCAATGATGCTGGGCGGCAATGTTGCGCGGCTCAAGAGCGATG 2884
Qy 1532 TAATGAAAGACATGCCCCCAAAAAAGAGAGCTCAATTTTGGGTGTTGATCTGAGCAGTCTGC 1591
Db 2885 TGTTCAGAAACATGCCCTCCAGACAGAACTAATTTGCGGTGGAGCTGAGCCCTATGC 2944
Qy 1592 AGAAAGAAATTAACAAGCTATTTTACCCGTTAATTTATCAAGTATTGA---CAAAAAAGG 1648
Db 2945 AGAAAGAAATACTACAAAGTACATCTCTCACTCGAAATTTTGAAGCACTCAATGCCCGAGGTG 3004
Qy 1649 GAGGTCTCAAAATTTCCCTTAATAACATTTATGATGAAATTTACGAAAGTATGTCGCCATC 1708
Db 3005 GTGGCAACAGGTTGCTCTGCTGAATGTGGTGAATGATCTTAAGAAAGTGTCTGAACCATC 3064
Qy 1709 CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTCACGACGCAAAATGAAGCTT 1759
Db 3065 CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCTTAAGATGCCCTAATGGCATGTATGATG 3124
Qy 1760 TCAACAACATTTTGGAGTCTTGGGAAAGCTGCAATCTTCTAGATAAATGATGGTCAAAAC 1819
Db 3125 GCAGTGCCTTAATCAGAGCATCTGGGAAATTTATTTGCTGTGCAAGAAATGCTCAAGAAC 3184
Qy 1820 TGAAGAGCAAGACACAGAGTCTTAATATACACAGATTTACAGCATATGCTGAGCTTAC 1879
Db 3185 TTAAGAGAGGTGGGCATCGTGTACTCATCTTTTCCAGATGACCAAGATGCTAGACCTGC 3244
Qy 1880 TTGAAGACTACTGTACCCCATTAAGAAATGGCAGTACGAGCGAAATTTGATGAAAGAGTTGGCG 1939
Db 3245 TAGAGAGTTTCTGGAAACATGAAGTTTAAATACGAACGATCGATGGTGGATCACTG 3304
Qy 1940 GAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAATCTTAACAAGTTTGT 1999

Db 3305 GGAACATGCGGCAAGAGGCCAATGACCGCTCAATGCACCGGCTCAGCAGTTCTGCT 3364
Qy 2000 TTTTGCTCTCCAAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA 2059
Db 3365 TCTTGCTTTCCACTCGAGCTGGGGGCTTGGGAATCAATCTGGCCACTGCTGACACAGTTA 3424
Qy 2060 TCATTTATGACAGTACTGGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
Db 3425 TTATCTATGACTCTGACTGGGAACCCCAATGATCAATTCAGGCTTTTAGCAGAGCTCACC 3484
Qy 2120 GACTTGGCCAAACAAATAAGTGATGATTTATAGGCTCATAAACCGAGGCACCAATTCGAAG 2179
Db 3485 GGATTTGGCAAAATAAAGAGTATGATCTACCGGTTGTGACCCGTCGCTCAGTGGAGG 3544
Qy 2180 AAAGGATGATGCAATTTGACTAAAGAAATAGTCTTAGAGCATCTTTGTTGGTGG----- 2234
Db 3545 AGCGCATCAGCAGGTGGCAAGAAAGAAATGATGCTGACGCATCTAGTGTGGGCGCTG 3604
Qy 2235 -----GAAACTCAAAACACAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290
Db 3605 GGCTGGGCTCAGACACTGGATCTATGTCCAAACAGGAGCTTTGATGATATCCTCAATTTG 3664
Qy 2291 GATCAAAAGGAGCTTTTGTAGTAGAAGATGATGAAGCAGGAAGTCTCGAAAAATTCATT 2350
Db 3665 GCACTGAGGAACATTTCAAGGATGAAGCACTGATGGAGGAGAGACAACAAGAGGGAG 3724
Qy 2351 ATGATGATGGGCTATAGACAAATGCTTGA-----TCGTGATCTCGTGAGGACAGG 2404
Db 3725 AAGATAGCAGTGTATCCACTACGATGATAGGCCATTGAACGGCTCTGACCGTAACC 3784
Qy 2405 AAGTCTCAGTGGATGATGAAGCAGAGATGATTTAAAGGCTTTCAGGTGCTTAATT 2464
Db 3785 AGGATGAGACTGAAGACACAGAAATTCAGGCGCATGAATGAATATTGAGCTCATTCAAAG 3844
Qy 2465 TTGAATATATAGATGAAAAATGAGCAGCAGCATTTAGAGGCACAGAGAGTCTGCTGAAA 2524
Db 3845 TGGCCCGATGTGTGACGGAAGAAAGAAATGGGGGAGGAAGAGAGTAGAAGCGGAAA 3904
Qy 2525 GCAAACTTTAGCAGGCGAATTTCTGATAGACAGATTTATGGGAAGAGTTGTTAAAGATA 2584
Db 3905 TCATTAAACAGGAAGAAGTGTGATCCTGA---CTACTGGGAGAAATTTGCTCGGCACC 3961
Qy 2585 AATTTGAGCTCCACGAGCTCAGGAGCTTAATGCTCTTGGAAGAAAGGAAGAGAGTCGCA 2644
Db 3962 ATTATGAGCAGCAGCAAGAAGATCTAGCCCAAACTTGGGCAAGGAAGAAGAAATCCGTA 4021
Qy 2645 AGCAGTTGGTATCCATTGAAG---AAGATGATCTTGTGTTTGGGAAGATGTGAGCTCTG 2701
Db 4022 AACAGGTCACACTACAATGATGGCTCCAGGAGGACCGAGATTGGCAGGACGACCACTCG 4081
Qy 2702 ATGGAGATGAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGGAGCAAGAGGTTTC 2761
Db 4082 ACAACCAAGTCCGATTTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTTGATGAACGTT 4141
Qy 2762 AGACGGTTCGACGCGCTACAGAAGAAAGGCTCGCGATAATTTTGGAAACCAA----- 2812
Db 4142 CAGAAGCTCCCGTAGGCCAGTCTGAAGGSCCTGCGGATGATAAAGATAGCCATTGC 4201
Qy 2813 CTCGTTGATGGAAGGTGAGGGGAGATCTTTTACAGATCTGGGTTTCAACAGAGTCAAA 2872
Db 4202 CTCCTCTGTTGGCCGCTGTTGGTGGGAATATTGAAGTACTTGGTTTTAAATGCTCGTCAGC 4261
Qy 2873 GGGCCATTTTGTACAGACTTTGATGAGGTATGG-----AGCTGGCNAATTTGATT 2923
Db 4262 GAAAAGCCTTTCTTAATGCAATATGCGATATGTTATGCTATGCTGCCACCTCAGGATGCTTTTACTA 4321
Qy 2924 GGAAGGAGTTGTTGCTCTCGCTTAAAGCAGAGACCTTTTGAAGAAATAAATGAATATGAA 2983
Db 4322 CCCAGTGGCTTGTAAAGACCTTCGAGGCAATCAGAGAAGAGTTCAAGGCATATGTCT 4381
Qy 2984 TACTCTTCTTGAAGCACATTCCTGAAGAAATAGACGAGAATTTCTCCAACTTTTCAGATG 3043

Db 4382 CTCTTTTCATCGCGCATTTATGTGAGCGGGGCGAGATGGGCTGAGACCTTTTGCTGATG 4441
Qy 3044 GTCTGCCCAAGGAGAGACTTAGATAGAGATGTTCTAGTCAGAATTTGCTCTTCTGATAC 3103
Db 4442 GTCTCCCGGAGAAGGCGCTGCTCGCCAGCATGCTCTTACTAGAAATGGTGTATGTCTT 4501
Qy 3104 TACTTCAGAGAAAGGTTGAAATTTGTAGAAGATCATCCAGGG 3144
Db 4502 TGATTCGCAAGAGGTTCAGGAGTTTGAACATGTTAATGGG 4542

RESULT 7
US-09-962-436-288
; Sequence 288, Application US/09962436
; Patent No. US20020081301A1
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-75
; CURRENT APPLICATION NUMBER: US/09/962,436
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,082
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/234,924
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 568
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 288
; LENGTH: 6417
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-962-436-288

Query Match 10.1%; Score 421.4; DB 9; Length 6417;
Best Local Similarity 51.9%; Pred. No. 3.7e-101;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;

Qy 818 TACATCCATACCAGCTTGAGGGACTTAATTTTTTGGGTTCTCGTGTCAAAAACAGACGC 877
Db 2266 TGCACCCCTATCAAAATGAGGCGCTTGAATGGTTGGCTTCTCTGGGCTCAGGCGACTG 2325
Qy 878 ATGTAATCTCTGCTGATGAAATGGGACTAGGCAAGACAATTCAAAGCATTCGCCCTTTTAG 937
Db 2326 ACACCATCTGGCTGATGAGATGGCTTGGGAAACTGTACAGACAGCAGTCTTCTGT 2385
Qy 938 CTTACATTTTTGAGGAG-----AACCTCATTCGGCATTTGGTAATTTGCTCTCTATCGA 991
Db 2386 ATTCCCTTTTACAAGGAGGGTCAATTCCAAAGGCCCTTCTCTAGTGAGCGCCCTCTTTCTA 2445
Qy 992 CTCTGCTTAACCTGGGAGAGAGTTTGGCCACATGGGCCCCACAGATGAAGCTGTTATGT 1051
Db 2446 CCATCATCAACTGGGAGCGGAGTTTGAATGTGGGCTCCAGACATGATGTGTAACCT 2505
Qy 1052 ATTTTGGCACTGGCGCAAGCTTCGAGCAGTTATTCAGAGAACAATGATTTTACTTATCGAAG 1111
Db 2506 ATGTGGGTGACAGGACAGCGGTGCCATCATCCGAGAGATGAGTTCTCTTTGAAGACA 2565
Qy 1112 ATCAAAAAAGATCAAGAAAAAATACTGGACAATAAGTAGTAGGAAAGCAAGCAAAAA 1171
Db 2566 ATGCCATTCGTGGTGGCAAGAGGCCCTC-----CCGCATGAAGAAAGAGGCAT 2613
Qy 1172 GAATCAAGTTTGTATGCTCTCCATCATGATGATGATCAACCTAGATTTCAGCAGTTC 1231
Db 2614 CTGTGAATTCATGTGCTGCTGACATCTTATGAATTTGATCACCATTGACATGGCTATTT 2673
Qy 1232 TAAAAACCAATTAAGTGGGAGTGCATGATTTGATGAAGGTTCATCGATAAAAAATAAGG 1291
Db 2674 TGGGCTCTATTGATTGGGCTGCTCATCGTGGATGAAGCCCATCGGCTCAGGAACAATC 2733
Qy 1292 ATTCAAGAGTGTCTCTTCTTATTGACACAGATTAATCAAGTACCAACCGTATTTCTTGTGACAG 1351
Db 2734 AGTCTAAGTTCTTCCGGGTATTGAATGGTTACTCACTCCAGCACAAGCTGTGTGCTGACTG 2793


```
Qy 2873 GGGCCATTTTGTACAGACTTTGTAGTATGG-----AGCTGGCAATTTTGATT 2923
Db 4351 GAAAGGCTTTCTTAATGCAATTATGCGATATGATGCCACCTCAGGATGCTTTTACTA 4410
Qy 2924 GGAAGGAGTTTGTCTCCTCGCTTAAAGCAGAGACCTTTTGAAGAAATAAATGAATATGAA 2983
Db 4411 CCCAGTGGCTTGTAAAGACCTCGGAGGCAATCAGAGAAGAGTTCCAGGCATATGTCT 4470
Qy 2984 TACTCTTCTTGAAGCACATTTGCTGAAGAAATAGACGAGAAATTCCTCAACCTTTTCAGATG 3043
Db 4471 CTCCTTTTCATGCGCATTTATGTCAGCCGGGGCAGATGGGCTGAGACCTTTGCTGATG 4530
Qy 3044 GTGTGCCAGGAGGAGCTTAGAATAGAAATGTTCTAGTCAGAAATGCTCTTCTGTATAC 3103
Db 4531 GTGTCCCGCAGAGAGGCTGTCTCGCCAGCATGTCCTTACTAGAAATTTGGTGTATGTCTT 4590
Qy 3104 TAGTTTCAGGAGAGGTGAAATTTCTAGAAATCATCCAGGG 3144
Db 4591 TGATTCGAAGAGGTTTCAGGAGTTTGAACATGTTAATGGG 4631
```

RESULT 9

US-10-037-270-325

; Sequence 325, Application US/10037270

; Publication No. US20030104529A1

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Liu, Chenghua

; APPLICANT: Asundi, Vinod

; APPLICANT: Zhang, Jie

; APPLICANT: Ren, Feiyan

; APPLICANT: Chen, Rui-hong

; APPLICANT: Zhao, Qing A.

; APPLICANT: Wehrman, Tom

; APPLICANT: Xue, Aidong J.

; APPLICANT: Yang, Yonghong

; APPLICANT: Wang, Jian-Rui

; APPLICANT: Zhou, Ping

; APPLICANT: Ma, Yunqing

; APPLICANT: Wang, Dunrui

; APPLICANT: Wang, Zhiwei

; APPLICANT: Tillinghast, John

; APPLICANT: Drmanac, Radoje T.

; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and

; FILE REFERENCE: Polypeptides

; CURRENT APPLICATION NUMBER: US/10/037,270

; CURRENT FILING DATE: 2002-01-04

; PRIOR APPLICATION NUMBER: 09/552,317

; PRIOR FILING DATE: 2000-04-25

; PRIOR APPLICATION NUMBER: 09/488,725

; PRIOR FILING DATE: 2000-01-21

; NUMBER OF SEQ ID NOS: 1104

; SOFTWARE: pc_FL_genes Version 1.0

; SEQ ID NO 325

; LENGTH: 6475

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (219)..(5957)

US-10-037-270-325

Query Match 10.1%; Score 421.4; DB 15; Length 6475;

Best Local Similarity 51.9%; Pred. No. 3.7e-101;

Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;

Qy 818 TACATCCATACCAGCTTGAGGAGCTTAATTTTTTGGCGGTTCTCGTGGTCAAAAACAGACGC 877

Db 2395 TGCACCCCTATCAATGGAGGCGCTGAATTTGGTTGGCTCTCTCTGGGCTCAGGGCACTG 2454

Qy 878 ATGTAATCTTGTGCTGAATAATGGGACTAGGCAAGACAAATTCAAAGCAATTCGCCCTTTTAG 937

```
Db 2455 ACACCATTTTGGCTGATGAGATGGGCTTGGGAAAACTGTACAGACAGAGTCTTCTGT 2514
Qy 938 CTTTCACTTTTGGAGGAG-----AACCTCATTCGCAATTTGGTAAATTTGCTCTCTATCGA 991
Db 2515 ATTCCCTTTACAAAGGAGGGTCATTCCAAAGGCCCTTCTTAGTGAGGCCCTCTTTCTA 2574
Qy 992 CTCGCGTAACTGGGAGAGAGAGTTTGGCCACATGGGCCCCACAGATGAAGCTGGTATGT 1051
Db 2575 CCATCATCACTGGGAGCGGGAGTTTGAATGTGGGCTCCAGACATGTATGTCTGAACCT 2634
Qy 1052 ATTTTGGCACTCGCGAAGCTCGAGCAGTTATCAGAGAAATGAGTTTACTTATCGAAAG 1111
Db 2635 ATGTGGTGACAAAGACAGCCGTGCCATCATCCGAGAGAAATGAGTTCTCTTTTGAAGACA 2694
Qy 1112 ATCAAAAAAAGATCAAGAAAAAAGAAATCTGGACAAATAAGTAGCGAAAAAGCAAAAAA 1171
Db 2695 ATGCCATTGCTGGTGGCAAGAGGCCTC-----CCGCATGAAGAAAGAGGCAT 2742
Qy 1172 GAATCAAGTTTGTATGTCTCTCCATCGTATGAGATGATCAACCTTAGATTCAGCAGTTC 1231
Db 2743 CTGTGAATTTCCATGTGCTGCTGACATCCTATGAATTTGATCACCATTGACATGGCTATTT 2802
Qy 1232 TAAACCAATTAAGTCGGAGTGCATGATTCTTCAATGAGGTCTACGACTCAAAAAATAAGG 1291
Db 2803 TGGGCTCTATTGATTGGGCTGCTCATCTGTGATGAAGCCCATCGGCTGAGAACATC 2862
Qy 1292 ATTCAAAAGCTGTTCTCTTTCATGACACAGTATTCAAGTAAACCAACCGTATTCTTCTGACAG 1351
Db 2863 AGTCTAAGTTCTTCGGGTATTGAATGTTACTCACTCCAGCACAAAGCTGTTGCTGACTG 2922
Qy 1352 GAACACACTTTCAGAAACAATTTGGATGAATTTTTCATGCTCATGCTATTTCTTGTATCGCG 1411
Db 2923 GGACACATTTACAAAAAATACTCGAAGAGTGTGTTTCACTGCTCAACTTTTCTCACCCCG 2982
Qy 1412 GGAAGTTTGAAGTTTGGAGGAGTTCAGAGGAGTTCAGAGGAGTTCAGAGTATTAATCAAGAGAGC 1471
Db 2983 AGAGTTTCCAAATTTGGAAGGTTTTTGGAGAGTGTGCTGACATTCAGAGAGAGGACC 3042
Qy 1472 AGATCTCAAGGTTTCACAAAATGTTGGCTCCACATTTTGTCTCAGAGGGTAAAAAAGAGC 1531
Db 3043 AGATAAAAAAAGTGCATGACATGCTGGGCGCGACATGTTGCGCGGCTCAAGAGCGATG 3102
Qy 1532 TAATGAAAGACATGCCCCCAAAAAGAGAGTCAATTTTGGTGTGTTGATCTGAGCAGTCTGC 1591
Db 3103 TGTTCAGAAACATGCCCTCCAGACAGAACTAATTTGCGTGTGGAGCTGAGCCCTATGC 3162
Qy 1592 AGAAGAAATATTACAAAGCTATTTTACCGGTAAATTTATCAAGTATTGA---CAAAAAAGG 1648
Db 3163 AGAAGAAATACTACAAAGTACATCCTCACTCGAAATTTTGAAGCACTCAATGCCCGAGGTG 3222
Qy 1649 GAGGTGCTCAAAATTTCCCTTAATAACATTTATGATGGAATTTACGAAAAGTATGCTGCCATC 1708
Db 3223 GTGGCAACAGGTGCTCTGCTGAATGTGTTGATGATCTTAAGAGTGTCTGCAACCATC 3282
Qy 1709 CTTA-----TATGCTAGAGGGTTTGAGCCAGTATTATTCAGAGCAAAATGAAGCTT 1759
Db 3283 CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCCTAAGATGCCCTAATGGCATGTATGATG 3342
Qy 1760 TCAAAACATTTTGGAGTCTTGTGGAAGCTGCAACTTCTAGATAAAATGATGTTCAAC 1819
Db 3343 GCAGTCCCTTAATCAGAGCAATCTGGGAAATTTATGCTGCTGCAGAAAAATGCTCAAGAAC 3402
Qy 1820 TGAAGAGCAAGGACACAGAGTCTTAATAATACACAGATTTTACGATATGCTGGACTTAC 1879
Db 3403 TTAAGAGGGTGGGATCGTGTACTCATCTTTTCCAGATGACCAAGATGTAGACCTGC 3462
Qy 1880 TTGAAGACTACTGTACCCATAGAAATGGCAGTACGAGGCAATTTGATGAAAGAGTTGGCG 1939
Db 3463 TAGAGGATTTCTGGAAACATGAAGTTATAAATACGAAACGATCGATGGTGAATCACTG 3522
Qy 1940 GAGCTGAGCGGCAAAATACGCATAGATCGGTTCAATGCCAAAAATTTCTAACAGTTTGT 1999
```

Db 3523 GGAACATCGCGCAAGAGGCCATTGACCGCTCAATGACCGGGTGTCTCAGCAGTTCTGCT 3582
Qy 2000 TTTTGTCTCTCACAGAGCTGGTGGCTTAGGAATAAATCTTGCAACGGCTGATACAGTAA 2059
Db 3583 TCITGGCTTCACATCGAGCTGGGGGCTTTGGAATCAATCTGGCCACTGCTGACACAGTTA 3642
Qy 2060 TCATTATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
Db 3643 TTATCTATGACTCTGACTGGNACCCCATATGACATTCAGGCCCTTTAGCAGAGCTCACC 3702
Qy 2120 GACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCCACCATGGAAG 2179
Db 3703 GGAITGGCAAAATAAAAGGTGATGATCTACCGGTTTGTGACCCGTCGCTCAGTGAGG 3762
Qy 2180 AAAGGATGATCAATTGACTTAAAGAAATGGTTCTTAGAGCATCTTTGTTGG----- 2234
Db 3763 AGCCGATCACCGAGTGGCAAGAGAAATGATGCTGACGCATCTAGTGGTGGGCGCTG 3822
Qy 2235 ----GAAACTCAAAACACAAACATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG 2290
Db 3823 GGCTGGGCTCAAGACTGGATCTATGTCCTCAACAGAGGCTTGATGATATCTCAAAATTG 3882
Qy 2291 GATCAAAAGGAGCTTTTGTCTAGTGAAGATGATGAAGCAGGAAAGCTCTGGAAAAATTCATT 2350
Db 3883 GCACTGAGAACTATTCAAGGATGAAGCCACTGATGGAGGAGGACAAACAAGAGGGAG 3942
Qy 2351 ATGATGATCGGCTATAGACAAATGCTTGA-----TCGTGATCTGTGGAGCGCAGG 2404
Db 3943 AAGATAGCAGTGTTATCCACTACGATGATGAAGGCCATTTGAACGGCTCTAGACCGTAACC 4002
Qy 2405 AAGTCTCAGTGGATGATGAAGAGAGATGATTTTAAAGGCTTTCAAAGTGGCTCAATT 2464
Db 4003 AGGATGAGACTGAAGACACAGAAATGTCAGGCGATGAATGATATTTGAGCTCATTCAG 4062
Qy 2465 TTGAATATATAGTGAATAAGGAGCAGCAGCAGATTTAGAGGCACAGAGAGTCGCTCTGAAA 2524
Db 4063 TGGCCAGTATGTGTACGGGAAGAAATGCGGGAGGAGAGAGAGTAGAAGCGGAAA 4122
Qy 2525 GCAATCTTCAGCAGGCAATCTCTAGACAGCAAGTTATGGAGAGAGTTGTTAAAGATA 2584
Db 4123 TCATTAAACAGGAAGAAGTGTGATCTGTA---CTACTGGGAGAAATTTGTCGGGACC 4179
Qy 2585 AATTGAGCTCCACGAGCTCAGGAGCTTAATGCTCTTGGAAAAAGGAAGAGTCCGA 2644
Db 4180 ATTATGACAGCAGCAGAGATCTAGCCCAATCTGGCAAGAGAAAGATCCGTA 4239
Qy 2645 AGCAGTTGGTATCCATGGAAG---AAGATGATCTGTGGTTTGGAAAGATGTAGCTCTG 2701
Db 4240 AACAGGTCAACTACAAATGATGGCTCCAGGAGGACCGAGATTGGCAGGACGACAGTCCG 4299
Qy 2702 ATGGAGATGAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACGAGGATTC 2761
Db 4300 ACAACAGTCCGATTTACTCAGTGGCTTCAGAGGAAGGTGATGAAGACTTTGATGAACGTT 4359
Qy 2762 AGACGGTTCGACGGCGGTACAGAGAAGAGGTTCGGATAATTTGGAACCAA----- 2812
Db 4360 CAGAGCTCCCGTAGGCGGAGTGTGAAGGCCCTCGCGAATCATAAAGATGACCATGC 4419
Qy 2813 CTCGGTTGATGAAGGTGAGGGAGATCTTTTACAGATCTGGGTTTCAACAGAGTCAAA 2872
Db 4420 CTCCTCTGTTGGCCGCTGTTGGTGGGAATATGAAGTACTTTGGTTTAAATGCTCGTCAGC 4479
Qy 2873 GGGCCATTTTGTACAGACTTTGTAGGATGAG-----AGCTGGCAATTTTGTATT 2923
Db 4480 GAAAGCCCTTTCTTAATGCAATTTATGCGATATGATGTCACCTCAGGATGCTTTTACTA 4539
Qy 2924 GGAAGGAGTTGTTCTCGCTTAAAGCAGAGACCTTTGGAAGAAATAAATGAATATGGA 2983
Db 4540 CCCAGTGGCTTGTAGAGACCTCGGAGCAATCAGAGAAAGATTCAAGGCATATGTCT 4599
Qy 2984 TACTTCTTGAAGCACAATGCTGAAGAAATAGACAGAAATTTCCAACTTTTTCAGATG 3043
Db 4600 CTCCTTTTCATGGGCATTTATGTGAGCGGGGCGAGATGGGGCTCAGACCTTTTGTGATG 4659

Qy 3044 GTGTGCCAAGGAAGGACTTAGAATAGAAGATGTTCTTAGTCAGAATTGCTCTTCTGATAC 3103
Db 4660 GTGTCCCCCGAAGAGCGCTGTCTCGCCAGCATGCTCTTACTAGAATTGGTTATGCTT 4719
Qy 3104 TAGTTCAGGAGAAAGGTGAAATTTGTAGAAGATCATCCAGG 3144
Db 4720 TGATTCGCAAGAGGTTTCAGGAGTTTGAACATGTTAATGGG 4760

RESULT 10

US-10-117-722-325
; Sequence 325, Application US/10117722
; Publication No. US20030219744A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2BCIP
; CURRENT APPLICATION NUMBER: US/10/117,722
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: dt_FL_genes Version 1.0
; SEQ ID NO 325
; LENGTH: 6475
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (219)..(5957)
US-10-117-722-325

Query Match 10.1%; Score 421.4; DB 16; Length 6475;
Best Local Similarity 51.9%; Pred. No. 3.7e-101;
Matches 1236; Conservative 0; Mismatches 1076; Indels 69; Gaps 10;

Qy 818 TACATCCATACCAAGCTTGAGGAGCTTAATTTTTTGGGGTCTCTCGTGGTCAAAACAGACGC 877
Db 2395 TGCACCCCTATCAATAGGAGGCGCTGAAATTTGGTTGGCTTCTCTGGGCTCAGGGCACTG 2454
Qy 878 ATGTAATCCTTGTGATGAATCGGACTAGGCAAGACAATTCAAAGCATTTGCCCTTTTAG 937
Db 2455 ACACCATCTTGGCTGATGATGGGCTTGGGAAAACCTGTACAGACAGAGTCTTCTGT 2514
Qy 938 CTTTCACTTTTGGAGGAG-----AACCTCATTTCCGATTTGGTAATTTGCTCTCTATCGA 991
Db 2515 ATTCCCTTTACAGAGGAGGTCATTCAAAGGCCCTTCTCTAGTAGGCCCTCTTCTTA 2574
Qy 992 CTCTGGCTAACTGGGAGAGAGATTGCGCACATCGGGCCCCACAGATGAAGCTGTTATGT 1051
Db 2575 CCATCATCAACTGGGAGCGGAGTTTGAATTTGGGCTCCAGACATGTATGTCTAAACCT 2634
Qy 1052 ATTTTGGCACTGGCGAGCTCGAGAGTATATCAGAGAAATAGTATGTTTACTTATCGAAAG 1111
Db 2635 ATGTGGTGAACAAGACAGCCGTGCCATCATCCGAGAGATGAGTCTCTCTTTGAAGACA 2694
Qy 1112 ATCAAAAAAAGATCAAGAAAAAGAAATCTGGACAAAATAAGTAGCGAAAGCAAGCAAAAA 1171
Db 2695 ATGCCATCTGTTGGTGGCAAGAGGCCCTC-----CCGATGAAGAAAGAGCAT 2742
Qy 1172 GAATCAAGATTTGATGCTCTCTCACATCGTATGAGATGATCAACCTAGATTCAGCAGTTC 1231
Db 2743 CTGTGAATTTCCATGTGCTGCTGACATCCTATGAAATTTGATCACCATTGACATGCGCTATTT 2802

1232	TAATAACCAATTAAAGTGGAGTGCATGATTGTTGATGAAGGTCAATCGACTGAAAAATAAGG	1291
2803	TGGGCTCTATTGATTGGGCGCTGCTCATCTGTGATGAAGCCCATCGCTGAAGACAATC	2862
1292	ATTCAAAGCTGTTCTCTTCAATTCACACAGATATTCAAAGTAACCAACCGTATTCTTCTGCAG	1351
2863	AGTCTAAGTTCTTTCCGGGTATTAAGATGGTTACTCACTCCAGACACAAGCTGTTGCTGACTG	2922
1352	GAACACACATTCAGAACCAACTCGATGAACTTTTTCATGCTCATGCAATTTTCTTCATGTCGG	1411
2923	GGACACCAATTACAAAACAATCTGGAAGATTGTTTCATCTGCTCAACTTCTCACCCCG	2982
1412	GGAAGTTTGAAGTTTGGAGGAGTTCAGGAGAGTTCCAGAGATTTCAAAGATATTTAATCAAGAGAGC	1471
2983	AGAAGTTTCCACAATTTGGAAAGTTTTTTTGAGGAGTTTGTGACATTTGCCAAGGAGGACC	3042
1472	AGATCTCAAGTTTGCACAAAATGTGGCTCCACATTTTGTCTCAGAAGGGTAAAAAAGAGC	1531
3043	AGATAAAAACAATGCGATGACATGCTGGGCGCGACATGTTGCGGCGCTCAAAGCCGATG	3102
1532	TAATGAAGACATGCCCCCAAAAAGAGAGCTCAATTTTTCGTGTGTGATCTGAGCAGTCTGC	1591
3103	TGTTCAAGACATGCCCCCAAGACAGAACTAATTTGTGCGTGTGGAGCTGAGCCCTATGC	3162
1592	AGAAAGAATATTAACAAGCTATTTTATACCGTAAATTATCAAAGTATTGA----CAAAAAGG	1648
3163	AGAAGAATACTACAAGTACATCTCACTCGAAATTTTGAAGCACTCAATTGCCTCGAGGTG	3222
1649	GAGTGCTCAAAATTTCCCTTTAATAACATTTATGATGGAATTCAGAAAAGTATGCTGCCATC	1708
3223	GTGGCAACCAAGGTGCTCTGCTGAAATGTGGTGAATGATCTTAAGAAGTGTCTGCAACCATC	3282
1709	CTTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTTCAGCGCAAAATCAAGCTT	1759
3283	CATACCTCTTCCCTGTGGCTGCAATGGAAGCTCTTAAGATGCCTTAATGGCATGTATGATG	3342
1760	TCAACAACATTTTGGAGTCTGTGGAAGCTGCAATTTCTAGATAAAATCATGGTCAAAAC	1819
3343	GCAAGTGCCTTAATCAGAGCATCTGGGAAATTTATTTGCTGCTGCAGAAAATGCTCAAGAACC	3402
1820	TGAAGAGCAAGGACACAGAGTCTTAATATACACACAGATTTTCAGCATATGCTCGACTTAC	1879
3403	TTAAGGAGGTGGGCATCTGTACTCATCTTTTCCAGATGACCAAGATGCTAGACCTGC	3462
1880	TTGAAGACTACTGTACCATAAGAAATGGCAGTACGAGCGAAATGTATGGAAGAGTTGGCG	1939
3463	TAGAGGATTTCTTGGAACTGAAGTTATAAATACGAACGCATCGATGTTGGTAATCACTG	3522
1940	GAGCTGAGCGGCAATACGCATAGATCGGTTCAATGCCCCAAAATTTCTAAAGATTTTGTT	1999
3523	GGAACTATGCGCGCAAGAGCCATTTACCGCTTCAATGCAACCGGGTGTCTCAGCAGTTCTGCT	3582
2000	TTTGTGCTCTCCAAAGAGCTGGTGGCTTAGGAATAAATCTTGCAAACGGCTGATACAGTAA	2059
3583	TC TTGCTTCCATCTGAGCTGGGGCCCTTGGAAATCAATCTGGCCACTGCTGACACAGTTA	3642
2060	TCATTTATGACGTACTGGAATCTCTCATGCTGATCTTTCAAAGCAATGGCTAGAGCTCATC	2119
3643	TTATCTATGACTCTGACTGGAAACCCCAATAATGACATTTCAAGGCTTTTGTAGCAGAGCTCACC	3702
2120	GACTTGGCCAAAACAATAAGGTGATGATTTATATAGGCTCATAAACCGAGGCCACCATTTGAAG	2179
3703	GGATTGGGCAAAATAAAAGGTAATGATCTACCGGTTTGTGACCCGTTGCGTCAAGTGAGG	3762
2180	AAAGGATGATGCAATTGACTAAAAAGAAAATGGTTCTTAGAGCATCTTTGTTGTTGG-----	2234
3763	AGCGCATACGCAAGTGGCAAGAAGAAATATGATGCTGACGCATCTAGTGTGGCGGCTG	3822
2235	----GAAACTCAAAACAACAAATTAATCAGGAAGAGTTAGATGACATCATCAGGTATG	2290
3823	GGCTGGGCTCCAAGACTTGGATCTATGTGCCAAACAGGAGCTTGATGATATCTCTCAAAATTG	3882

RESULT 11

US-10-175-523-51

US-10-173-323-31
; Sequence 51, Application US/10175523

; Publication No. US20030096264A1

; GENERAL INFORMATION:

APPLICANT: Brockman, Jeffrey

APPLICANT: Evans, David

APPLICANT: Hook, Derek

APPLICANT: Klimczak, Leszek

APPLICANT: Laeng, Pascal

APPLICANT: Palfreyman, Michael

APPLICANT: Rajan, Prithi


```

; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/1J7995-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51
; LENGTH: 6331
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-175-523-51

```

[illegible]

```
Db 4131 CAAGCAGGAGGAGAAATGTGGACCTTGAC-----TACTGGGAGAGCT 4172
Qy 2574 GTTAAAGATAAATTTGAGCTGCAACAGGCTGAGAGCTTAATGCTCTTGGAAAGAGAA 2633
Db 4173 GCTGAGGCAATCACTATGAGCAACAGCAGGAAGACCTAGCCCGGAATCTAGGCAAGGGCAA 4232
Qy 2634 GAGAAGTCGCAAGCAGTTGGTATCCATTTGAAGAAGATGATCTTCTGTTTGGGAAGATGT 2693
Db 4233 GCGGGTTCGCAAGCAAGTTAACTAATATGATGCTGCTCAGAGAGACCAGAACACCAAGTC 4292
Qy 2694 GAGCTCTGATGGAGATGAAGATTTATGAAGCTGAGTCAACAGATGGTGAAGCAGCAGGACA 2753
Db 4293 AGAGTACTCGTGGTTCAGAGGAGGAGGATGAAGACTTCGATGAACGCTCTGAAGGGCG 4352
Qy 2754 AGGAGTTCAGACGGGTGACGCGCGTACAGAAAGAGGTGCGCATTAATTTGGAAACCAAC 2813
Db 4353 TAGACAGTCAAGAGGAGGAGCTCCGGAA-----TGAGAAAGATAAGCCACTGCC 4400
Qy 2814 TCCGTTGATGAAGGTGAGGGGAGATCTTTCAGAGTACTGGGTTTCAACCAAGACTCAAG 2873
Db 4401 TCCACTGCTGCCGAGTCGGGGGCAACATTTGAGGTGCTGGGCTTCAACACCCGTCAGCG 4460
Qy 2874 GGCCATTTTGTACAGACTTTGATGAGGTATGG-----AGCTGGCAATTTTGATG 2924
Db 4461 GAAGCTTTCCTCAATGCTGTATGCGCTGGGGATGCCACCACAGGATGCCCTTCAACCAC 4520
Qy 2925 GAAGGAGTTTGTCTCGCTTAAAGCAGAGACCTTTGAAGAAATAAATGAATATGAAT 2984
Db 4521 ACAGTGGCTGTGCGGGACCTGAGGGGCAAGACTGAGAGGAGTTTAAAGGCTATGTGTC 4580
Qy 2985 ACTCTTCTGAAGCACATTCGTGAAGAAATAGACGAGAAATCTCCAACTTTTCAGATGG 3044
Db 4581 TTTGTTTATGCGCCATCTGTGTAGCGCTGGGGCAGACGGCTCTGAAACCTTTGCGCGATGG 4640
Qy 3045 TGTGCCCCAAGGAAGGACTTAGAATAGAAAGATGTTCTAGTCAGAAATTCCTCTCTGTACT 3104
Db 4641 GGTCCCTCGGAGGAGCTGAGTCCCGCAGCGTGTTCACCCGATTCGAGTCATGTCTCT 4700
Qy 3105 AGTTCAGAGAGGTGAA 3122
Db 4701 CGTCAAAAAGAGGTGCA 4718

RESULT 12
US-10-159-563-176
; Sequence 176, Application US/10159563
; Publication No. US200400091541
; GENERAL INFORMATION:
; APPLICANT: Khan, Javed
; APPLICANT: Ringner, Markus
; APPLICANT: Peterson, Carsten
; APPLICANT: Meltzer, Paul
; TITLE OF INVENTION: SELECTIONS OF GENES AND METHODS OF USING THE SAME FOR
; TITLE OF INVENTION: DIAGNOSIS AND FOR TARGETING THE THERAPY OF SELECT CANCERS
; FILE REFERENCE: 11613.56US11
; CURRENT APPLICATION NUMBER: US/10/159,563
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US 10/133,937
; PRIOR FILING DATE: 2002-04-25
; NUMBER OF SEQ ID NOS: 444
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 176
; LENGTH: 6331
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-159-563-176
```

```
Query Match 10.1%; Score 420.8; DB 16; Length 6331;
Best Local Similarity 52.3%; Pred. No. 5.3e-101;
Matches 1233; Conservative 0; Mismatches 1032; Indels 93; Gaps 10;
Qy 810 AGGCTTTTACATCCATACAGCTTTGAGGACCTTAATTTTGGGTTCTCGTGGTCAAA 869
|||||
```

```
Db 2409 AGGCAACCTGCACATGATATCAGTTGGNAGGCTGAACTGGCTAGGCTTCTCCTGGGCCCA 2468
Qy 870 ACAGAGCATGTAATCTCTGCTGATGAAATGGAGCTAGGCAAGACAATTCAAAGCATGTC 929
Db 2469 GGGCACTGACACCATTTCTAGCTGATGAGATGGGGCTAGGCAAGACCATACAAACCATCGT 2528
Qy 930 CCTTTTAGCTTCACTTTTGGAGGAGAACCTCA-----TTCCGCATTTGGTAAATGCTCC 983
Db 2529 CTTCTCTTACTACTCTTCAAGGAGGGGCCACACAAAAGGTCCCTTCTCTGTTGATGGCCCC 2588
Qy 984 TCTATCGACTCTCGTAACTGGGAGAGAGAGTTTGGCACATGGGCCCCCACAGATGAACGT 1043
Db 2589 ACTCTCTACCATCATTTAACTGGGAGCGGAGTTCAGATGTCGGCACCCCAATTTCTATGT 2648
Qy 1044 GGTATGATTTTGGCACTGCGCAAGCTCGAGCAGTTATCAGAGAACATGAGTTTACTTT 1103
Db 2649 GGTGACATACACGGGTGACAAAGGACAGCGGGCCATCATTCGTGAGAAATGAATTTCTCTT 2708
Qy 1104 ATCGAAGATCAAAAAAGATCAAGAAAGAAATCTGGACAATTAAGTAGCGNAAGCAA 1163
Db 2709 TGAGGACAAATGCCATCAAGGGGGCAAGAAAGCTTTTAAAGATGAAAAG----- 2756
Qy 1164 GCAAAAAAGAAATCAAGTTTGTATGCTCCTCCTCACATGCTATGAGATGATCAACCTAGATTC 1223
Db 2757 GGAGGCACAGGTGAAGTTCATGTTCTCTGACATGCTATGAGCTGATCACCATTGATCA 2816
Qy 1224 AGCAGTTCTAAACCATAAAGTGGAGTGCATGATTTGTATGAAGGTCAATCGACTGAA 1283
Db 2817 GGCAGCACTTGGTTCCATCGCTGGGCCCTGCTCTGTGTGATAGTAGGCCCCATCGACTCAA 2876
Qy 1284 AATAAGGATTCAAAAGCTGTTCTCTCATTTGACACAGATATTCAAGTAAACCACGTTACT 1343
Db 2877 GAACAACCAAGTCCAAAGTTTTCAGGGTTCTCAATGGTTTCAAGATAGATCATAAAGTTGCT 2936
Qy 1344 TCTGACAGGAACCACTTTCAGAAACAATTTGAGTGAATTTTTCATGCTCATGCAATTTCT 1403
Db 2937 GCTGACAGAAACCCATTGAGANATTAATCTGGAGGAGCTCTTCCATCTCTGAACTTCT 2996
Qy 1404 TGATCGGGGAGTTTGGAAAGTTTGGAGGTTTCCAGGAGGTTTCAAGAGATTAATATCA 1463
Db 2997 CACCCACAGAGAGATTTAAACAACCTTGGAGGGCTTCTCTGGAGGAGTTTGTGACATATCAA 3056
Qy 1464 AGAGGAGCAGATCTCAAGGTTCACAAAATGTTGGTCCACATTTGCTCAGAGGGTAAA 1523
Db 3057 AGAGGACAGATCAAGAAACTGCAATGTTGCTGGGGCCACACATGCTCGGAGACTCAA 3116
Qy 1524 AAAGAGCTAATGAAAGACATGCCCCCAAAAGGAGCTCATTTTGGCTGTTGATCTGAG 1583
Db 3117 GGCAGATGCTTTAAGAACATGCCACCACAGACAGAGCTCATGTTGGGTGAGGCTAAG 3176
Qy 1584 CAGTCTGCAGAAAGAAATATTACAAAGCTATTTTACCCTGTAATTAATCAAGTATTGA---C 1640
Db 3177 CCCATGCAAGAGAAATACTACAAATACATCTGACTCGAAATTTTGGAGCCCTTGAATTC 3236
Qy 1641 AAAAAGGAGGCTCAAAATTTCCCTTAATACATTAATGATGAATTAAGCAAGATG 1700
Db 3237 ACAGGTGGTGGGAACCAAGTGTGCTGCTTAATATCATGATGGATCTTTAAGAAGTGTG 3296
Qy 1701 CTGCCATCTTATATGCTAGAGGGTGTG-----AGCCAGTTTATTCAACGACCAA 1751
Db 3297 CAACCATCACTACTTTTCCCGTGGCTGCTATGGAGTCCCCCAAACTCCCAAGTGGGC 3356
Qy 1752 TGAAGCTTTCAACAACTTTTGGAGTCTTTGTGGAAAGCTGCAACTTTCTAGATAAAATGAT 1811
Db 3357 TTATGAGGGTGGGGCACTTATTAAAGTCGTCGGGAAGCTCATGCTCTCCAGAAGATGCT 3416
Qy 1812 GGTCAAACTGAAGAGCAAGGACACAGAGTCTTAATATACACACAGATTTTCAGCATATGCT 1871
Db 3417 GCGAAAGCTGAAGGAGCAAGGACACCGAGTGTCTCTTCTCGCAGATGACCAAAATGTT 3476
Qy 1872 GGACTTTACTTTGAGACTACTGTACCATAGAAATGCGCAGTAGCGGCAATTTGATGAAA 1931
Db 3477 AGACTTGTCTGAGGACTTCTTTAGACTATGAAGGCTACAAGATGAGCGCATTCATGTTGGTGG 3536
```


Db 3461 AGAACACCAGTCCAAAGTTTTCAGGGTCTCAATGGTTACAGATAGATCAATAGTTGC 3402
Qy 1343 TTCTGACAGGAACACCACTTTCAGAAACAACTTGGATGAACCTTTTCATGCTCATGCAATTTTC 1402
Db 3401 TGCTGACAGGAACCCCATTTGCAGAAATATCGAGGAGCTCTTCCATCTCCTGAACTTC 3342
Qy 1403 TTGATGGGGGAGTTTGGAGTTTGGAGGTTCCAGGAGGAGTTCAAGATATTAATC 1462
Db 3341 TCACCCAGAGAGATTTTAAACAATTTGAGGGCTTCTCGGAGGATTTGCTGACATATCCA 3282
Qy 1463 AAGAGGAGCAGATCTCAAGTTTGACAAATTTGGCTCCACATTTGCTCAGAAAGGTAA 1522
Db 3281 AAGAGGACAGATCAAGAACTGCATGATTTGCTGGGGCCACACATGCTCGGAGACTCA 3222
Qy 1523 AAAAAGCAGTAATGAAGAAGATGCCCCCAAAAGGAGCTCATTTTGGCGTTGATCTGA 1582
Db 3221 AGGCAGATGTCTTTAAGAACATGCCAGCAAGACAGAGCTCATCTGTCGGGTGGAGCTAA 3162
Qy 1583 GCAGTCTGCAGAAAGATATTACAAAGCTATTTTACCCGTAATTTATCAAGTATTGA--- 1639
Db 3161 GCCCATTCGAGAGAAATATCTACAAATACATCTGACTCGAAATTTTGGGCGCTTGAAT 3102
Qy 1640 CAAAAAGGAGGTGCTCAAAATTTCCCTTAATAACATTATGATGGAATTAACGAAAGTAT 1699
Db 3101 CACGAGTGTGGAAACAGGTGCTGCTGCTTAATATCAATGATGATCTTAAAGATGCT 3042
Qy 1700 GCTGCCATCTCTTATGCTAGAGGGTGTG-----AGCCAGTTTATTCAGGACGCAA 1750
Db 3041 GCAACCATCCATACCTTTTCCCGTGTCTGATGGAGTCCCCCAAACTCCCCAGTGGG 2982
Qy 1751 ATGAGCTTTCAAAACAATTTTGGAGTCTTGTGGAAGCTGCAACTTCTAGATAAATGA 1810
Db 2981 CTTATGAGGTGGGGCACTTTAAGTCTGTGGAAGCTCATGCTGCTCCAGAAATGC 2922
Qy 1811 TGGTCAAACTGAAAGACAGGACACAGAGTCTCTAATATACACACAGTTTTCAGCATATGC 1870
Db 2921 TGGAAAGCTGAAGGACAAAGGACACCGAGTGTCTCTTCTCGAGATGACCAAAATGT 2862
Qy 1871 TGACATTTACTTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGAATTTGATGGAA 1930
Db 2861 TAGACTTGTCTGAGGACTCTTAGACTATAGGCTACAGTATGAGCGCATCGATGGTG 2802
Qy 1931 AGGTGGCGGAGCTGAGGCGAAATACGATAGATCGGTTCAATGCCAAAAATTTCTAACA 1990
Db 2801 GTATCACGGTGCCTGAGGACGAGGCCATCGATCGGTTTAATGTCTCTGGGGCCCAAC 2742
Qy 1991 AGTTTGTGTTTGTCTCTCACAGAGCTGTTGGCTTAGGAATAAATCTTGCAACGCTG 2050
Db 2741 AATTCTGCTTCTCTCTGTCCACCGAGCTGGGGCCTGGGCAATCAATCTGGCCACTGCTG 2682
Qy 2051 ATACAGTAATCATTTATGACAGTGACTGGAAATCTCATGTGATCTTCAAGCAATGGCTA 2110
Db 2681 ACATGTTCATCATCTTTGATTTCTGACTGGAAACCCCAATATGATCCAGGCCCTTAGCC 2622
Qy 2111 GAGCTATGAGCTTGGCCAAACAAATAAGTGTATTTATAGCTCATATAACCGAGGCA 2170
Db 2621 GGGCTCATCGATTTGGCCAGGCCAAACAAAGTGTATTTACCGGTTTGTGACTCGCGCT 2562
Qy 2171 CCATTGAGAAAGGATGATGCAATTTGACTTAAAGAAATGGTTCTAGAGCATCTTTGTTG 2230
Db 2561 CAGTGGAAAGCGCAATCACAAAGTGGCCAAAGAAAGATGATGCTGACACCTCTGGTTG 2502
Qy 2231 TTGG-----GAAACTCAAAACACAAAACATTAATCAGGAAGAGTTAGATGACATCA 2281
Db 2501 TGGCGCTGGGCTCCAGCAGGCTCCATGTCCAGCAGGAGCTTTCAGCATTC 2442
Qy 2282 TCAGGTATGATCAAGAGAGCTTTTGTCTAGTGAAGATGATGAAGCAGGAAAGCTG--- 2338
Db 2441 TCAAATTTGGCACTGAAGAGCTATTCAAGGATGAAACGAGGGGAGAAACAAGGAGAGG 2382
Qy 2339 -----GAAAAATTCATTATGATGCGGCTATAGACAAATTTGCTTGTATCGTATCTCG 2392

Db 2381 ACAGCAGTGTGATTCATTATGACAAATGAGGCCATCGCTCGGCTGTTGGACCGGAACCAAG 2322
Qy 2393 TGGAGCCAGAGAGAGTCTCAGTGGATGATGAGAGAGAAATGGAATCTTTAAAGGCTTCA 2452
Db 2321 ATGCAACTGAGG-----CACTGACGTGCAGAAATGATGATCTCAGCTCCTTCA 2268
Qy 2453 AGGTGCTTAATTTTGAATATATAGATGAAAATGAGGCAGCAGCATTAGAGGCACAGAG 2512
Db 2267 AGTGGCAAGTACGTCGTCGCGGGAAGAAGACAAGATTGAGGAAATTTAGCAGAGATCA 2208
Qy 2513 TGGTCTCTGAAAGCAAAATCTTCAGCAGGCAATTTCTGATAGCAAGTTTATTTGGAAAGAT 2572
Db 2207 TCAAGCAGGAGGAGATGTGGACCTGAC-----TACTGGGAGAGC 2166
Qy 2573 TGTAAAAAGATAAATTTGAGCTGCACCGCTGAGGAGCTTAATGCTCTTTGGAAAAAGGA 2632
Db 2165 TGCTGAGGCATCACTATGAGCAACAGCAGGAAGACCTAGCCCGGAATCTAGGCAAGGGCA 2106
Qy 2633 AGAAGAGTCGACAGCAGTTGGTATCCATTTGAAGAGATGATCTTGTGTTTGGAAAGATG 2692
Db 2105 AGCGGTTTCGCAAGCAAGTTTAACTACAATGATGCTGCTCAGGAAGACCAAGACAACAGT 2046
Qy 2693 TGAGCTCTGATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGTTGAAGCAGCAGGAC 2752
Db 2045 CAGAGTACTCGTGGGTTTCAGAGGAGGAGATGAAGACTTCGATGAACGCTCTGAAGGGC 1986
Qy 2753 AAGAGGTTTCAGACGGGTCGACCGCCGTACAGAAAGAGGTCGCGATAATTTGGAAACCAA 2812
Db 1985 GTAGACAGTCAAGAGGAGCTCCGGAA-----TGAGAAAGATTAAGCCACTGC 1938
Qy 2813 CTCGCTTGATGAAAGTGAAGGGGAGATCTTTCAGAGTACTGGGTTTCAACAGAGTCAAA 2872
Db 1937 CTCCACTGTGTCGCCGAGTCGGGGGCAACATTTGAGGTGCTGGGCTTCAACACCCGTCAG 1878
Qy 2873 GGCCCAATTTGTACAGACTTTTGTAGGTATGG-----AGCTGGCAATTTTGATT 2923
Db 1877 GGAAGCTTTCTCAATGCTGTGATGCGCTGGGGATGCCACACAGGATGCCTTCAACA 1818
Qy 2924 GGAAGAGTTTGTCTCGCTTAAAGCAGAAAGACCTTTGAAGAAATTAATGAAATATGGAA 2983
Db 1817 CACAGTGGCTGTGTCGGGACCTGAGGGCAAGACTGAGAAGGAGTTTAAAGGCTATGTGT 1758
Qy 2984 TACTCTTTTGAAGCNCATTTGCTGAAGAAATAGACGAGAAATTTCTCAACCTTTTCAGATG 3043
Db 1757 CTTTGTTCATGCGCCATCTGTGTGAGCCTGGGGCAGACGGCTCTGAAACCTTTGCGCATG 1698
Qy 3044 GTGTGCCCAAGAAAGGACTTAGAATAAGAGATGTTCTAGTCAGAAATTTGCTTCTTCATAC 3103
Db 1697 GGGTCCCTCGGAGGAGCTGAGTCCGACAGGTTGTGACCGCATTTGAGTCATGTCTC 1638
Qy 3104 TAGTTTCAGGAGAGGTGAA 3122
Db 1637 TCGTCAAAAAGAGGTGCA 1619

RESULT 14

US-10-044-090-370
; Sequence 370, Application US/10044090
; Publication No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 370
; LENGTH: 7805
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

; OTHER INFORMATION: Incyte ID No. US20020137081A1 235885.5
US-10-044-090-370

Query Match	9.8%;	Score 409.6;	DB 14;	Length 7805;
Best Local Similarity	52.5%;	Pred. No. 6.2e-98;		
Matches 1240;	Conservative 0;	Mismatches 1024;	Indels 96;	Gaps 12;

Qy	810	AGGCTGTGTTACATCCATACACAGCTTGGAGGACTTAAATTTTTTGGGGTTCCTCGTGGTCAAA	869
Db	2692	AGGCACCTGGACATGATGATGTTGGAGGCTGAACTGGCTAGCTTCTCTGGGCCCA	2751
Qy	870	ACAGACGCATGTAATCTCTGATGAAATGGGACTAGGCAAGCAAAATCAAAGCATGTC	929
Db	2752	GGGCACCTGACACCATCTAGCTGATGATGAGTGGGCTAGGCAAGCAACCATCAAAACCATCGT	2811
Qy	930	CTTTTAGCTTCACTTTTGGAGGAACTCA-----TTCCGCAATTTGGTAATGCTCC	983
Db	2812	CTTCCTCTACTCACTCTACAAGAGGGGCCACAAAAAGGTCCTCTCGTGGTGGTGGCCC	2871
Qy	984	TCATTCGACTCTGCTTAACCTGGGAGAGAGTGTTCACACATGGGCCCCACACAGATGAACGT	1043
Db	2872	ACTCTTACCATCATTAACCTGGAGCGGAGTTCAGATGTGGCACCCAAATTTCTATGT	2931
Qy	1044	GGTATGATTTTGGCACTGCGCAAGCTCGAGCAGTTATACAGAAATGAGTTTACTT	1103
Db	2932	GGTGACATACACGGGTGACAAAGGACAGCCGGGCCATCATTCGTGAGAATGAATTTCTCCTT	2991
Qy	1104	ATCGAAGAT-CAAAAAGAGTCAAGAAAAGAAATCTGGCAAAATAGTAGCAAAAGCA	1162
Db	2992	TGAGGACATGCCATCAAGAGGGGCCAAGAAAGCTTTTAAAGATGAAGAGGGGACACAGG-	3050
Qy	1163	AGCAAAAAGAAATCAAGTTTCACTGCTCTCCATCATCGTATGAGATGATCAACCTAGATT	1222
Db	3051	-----TGAGTTTCCATGTTCTCTGACATCGTATGAGCTGATCACCATTGATC	3098
Qy	1223	CAGCAGTTCTAAAAACCAATTAAGTGGGAGTGCATGATTTGTTGATGAAGGTCATCGACTGA	1282
Db	3099	AGGCAGCACTGGTTCCATCCGCTGGGCTGTCTTGGTAGATGAGGCCCATCGACTCA	3158
Qy	1283	AAATAGGATCAAGCTGTTCTTCTCATTTGACACACAGTATTCAGTAACACCGTATTC	1342
Db	3159	AGAACACACAGTCCAAAGTTTTTCAGGGTTCTCAATGGTTACAAGATAGATCATAGTTGC	3218
Qy	1343	TTCTGACAGGAACACCACTTCAGAACCACTTGGATGAATTTTCATGCTCATGCAATTTTC	1402
Db	3219	TGCTGACAGGAACCCCATTCAGAAATATCTGGAGGAGCTTCCATCTCTGAACTTCC	3278
Qy	1403	TTGATGCGGGGAAGTTTGGAGTTGGAGGAGTTCAGAGGAGTTCCAGAGGATTAATATC	1462
Db	3279	TCACCCAGAGAGATTTAAACCACTTGGAGGGCTTCTGTGGAGGAGTTTGTCTGACATATCCA	3338
Qy	1463	AAGAGGACAGATCTCAAGTTTGCACAAATGTTGGCTCCACATTTGCTCAGAGGGTAA	1522
Db	3339	AAGAGGCCAGATCAAGAAATCTGATGATTTGCTGGGGCCACACATGCTCGGAGACTCA	3398
Qy	1523	AAAAAGACGTAATGAAACACATGCCCCCAAAAAGGAGCTCATTTTGGCTGTGATCTGA	1582
Db	3399	AGGCAGATGCTTTAAGAACATGCCAGCAAGACAGAGCTCATCGTTGGGTGGAGCTAA	3458
Qy	1583	GCAGTCTCCAGAAAGAAATATTACAAAGCTATTTTACCCGPAATATCAAGATATGA---	1639
Db	3459	GCCCCATGCAGAAAGAAATACTACAAATACATCTGACTCGAAATTTTGGAGCCCTTGAAT	3518
Qy	1640	CAAAAAGGGAGGTGCTCAAAATTTCCCTTAATACATTTATGATGAAATACGAAAGTAT	1699
Db	3519	CACGAGGTGGTGGGAACACAGGTGTCGCTGCTTAAATATCATGATGATCTTAAAGAGTGTCT	3578
Qy	1700	GCTGCATCTTATATGCTAGAGGTGTG-----AGCCAGTTATTTCACACGCA	1750
Db	3579	GCAACCATCCATACCTTTTCCCGTGGCTGTATGGAGTCCCCCAACTCCCGAGTGGGG	3638
Qy	1751	ATGAGCTTTTCAAAACAACTTTTGGAGTCTTGTGGAAAGCTGCAACTTCTAGATAAAATGA	1810

Db	3639	CTTATCAGGGTGGGGCACTTTATTAACTGCTCTGGGAAGCTCATGCTGCTCCAGAAAGATGC	3698
Qy	1811	TGCTCAAACTGAAAGAGCAAGGACACAGAGTCCCTAATATATACACACAGTTTTCAGCATATGC	1870
Db	3699	TGCAAAAGCTGAAGGAGCAAGGACACGAGTGTCTCATCTTCTCGCAGATGACCAAAATGT	3758
Qy	1871	TGACCTTACTTGAAGACTACTGTACCCATAAGAAATGGCAGTACGAGCGAATTTGATGGAA	1930
Db	3759	TAGACTTGTCTGAGGACTCTTAGACTATGAAGGCTACAAGTATGAGCGCATCGATGGTG	3818
Qy	1931	AGTTGGCGAGCTGACGGCAAAATACGATAGATCGGTTCAATGCCAAAAATTTCTAACA	1990
Db	3819	GTATCACGGTGCGCTGAGCGCAGGCGCCATCGATCGGTTTAAATGCTCTGGGGCCCAAC	3878
Qy	1991	AGTTTGTGTTTTGCTCTCCACAAGAGCTGGTGGCTTAGGAAATAATCTTGGCAACGGCTG	2050
Db	3879	AAITCTGCTTCTCTGTCCACCCGAGCTGGGGCCCTGGGCATCAATCTGGCCACTGCTG	3938
Qy	2051	ATACAGTAATCATTTATGACAGTGAATCCTCATGCTGATCTTCAAGCAATGGCTA	2110
Db	3939	ACACTGTTCATCATCTTTGATTCTGACTGGAAACCCCATATGACATCCAGGCCCTTAGCC	3998
Qy	2111	GAGCTCATC-GACTTGGCCAAACAAATAGGTGATGATTTATAGGCTCATAAACCGAGGC	2169
Db	3999	GGCTCATCGGATTTGGGCCAGGCCAACAAAGTGAATTTTACCGGTTTGTGACTCGCGCG	4058
Qy	2170	ACCAATTGAAGAAAGGATGATGCAATTGACTTAAAGAAAAATGGTCTTAGAGCATCTTTGT	2229
Db	4059	TCAGTGAAGAGCGAATCACAAAGTGGCCAGAGAAAGATGATGCTGACACACCTGGTT	4118
Qy	2230	GTTGG-----GAAACTCAAAAACAAAAATTAAATCAGGAAGTGTAGATGACATC	2280
Db	4119	GTGCGGCTGGGCTGAGCTCCAAAGGAGGCTCCATGATCCAAAGCAGGAGCTTGACGACATT	4178
Qy	2281	ATCAGGTATGATCAAGAGGCTTTTGTGCTAGTGAAGATGATGAACGAGGAAGTCTG--	2338
Db	4179	CTCAAAATTTGGCACTGAAGAGCTATTCAAGGATGAAAAACGAGGGGAGAAACAAGAGGAG	4238
Qy	2339	-----GAAAAATTCATTATGATGATGCGGCTATAGACAAATTTGCTTGTATCGTGATCTC	2391
Db	4239	GACAGCGTGTGATTCATTATGACATGAGCCCATCGCTCGGCTGTTGGACCGGACCCAG	4298
Qy	2392	GTGAGGCGACAGAAAGTCTCAGTGGATGATGAAGAGAGAAATGGAATCTTTAAAGGCTTTC	2451
Db	4299	GATGCAACTGAGCA-----CACTGACGTGCAGAAACATGAATGATATCTCAGCTCCTTC	4352
Qy	2452	AAGGTGGCTAAATTTTGAATATATAGTGAATAATGAGCAGCAGCATTTAGAGGACACAGAG	2511
Db	4353	AAGGTGGCACAGTACGTCGTCGGGGAAGAGACAAGATTGAGGAAATTTAGCGGAGATC	4412
Qy	2512	GTGCTGCTGAAAAGCAAAATCTTCAGCAGGCAATTTCTGATAGAGCAAGTTATTGGGAAGAG	2571
Db	4413	ATCAAGCAGAGGAGATGTTGGACCTGAC-----TACTGGGAGAAG	4454
Qy	2572	TTGTTAAAAAGATAAATTTGAGCTGCACAGGCTGAGAGCTTAAATGCTCTTTGAAAAAGG	2631
Db	4455	CTGCTGAGGACATCACTATGAGCAACAGCAGAGAAACCTAGCCCGGAATCTAGGCAAGGCG	4514
Qy	2632	AAGAAAGTCCAGCAGGTTGGTATCCATTGAAGAAGATGATCTTGTGCTGTTTGGAAAGAT	2691
Db	4515	AAGCGGTTTCCGACGCAAGTTAACTACAATGATGCTCTCAGGAAGAACCAAGACACACAG	4574
Qy	2692	GTGAGCTCTGATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGTTGAAGCAGCAGGA	2751
Db	4575	TCAGAGTACTCGTGGGTTTCAAGGAGGAGGATGAAGACTTCGATGAACCTCTGNAAGGG	4634
Qy	2752	CAAGGAGTTCAGACGGGTCCAGCGGCCGTA CAGAGAAAGGGTCCGGTAATTTTGGAAACCA	2811
Db	4635	CGTAGACAGTCAAAAGAGGCGAGCTCCGGAA-----TGAGAAAGATGAAGCCACTG	4682
Qy	2812	ACTCCGTTGATGAGAGTGGAGGAGATCTTTTCAAGAGTACTGGGTTTTCACACAGAGTCAA	2871
Db	4683	CCTCCACTGTGCGGCCCGAGTCGGGGGCAACATTGAGGTGCTTGGGCTTCAACACCCGTCAG	4742

Qy	2872	AGGCCAATTTTGTACAGACCTTTGATGAGGTATGG-----AGCTGCAATTTTGTAT	2922
Db	4743	CGGAAGGCTTTTCTCAATGCTGTGATCGCTGGGGGATGCCACACAGGATGCCCTTCACC	4802
Qy	2923	TGGAAGGAGTTTGTTCCTCGCTTAAAGCAGAAGACCTTTTGAAGAAATAAATGAATATGGA	2982
Db	4803	ACACAGTGGCTGTGGCGGACCTGAGGGGCAAGNCTGAGAAGGNGTTTAAAGGCTATGTC	4862
Qy	2983	ATACTCTTCTTGAAGCACATTTGCTGAAGAAATAGACGAGAAATTCTCCAACTTTTCAGAT	3042
Db	4863	TCTTTGTTCATCGGCCATCTGTGTGAGCCTGGGGCAGCGCTCTGAAACCTTTGCCGAT	4922
Qy	3043	GGTGTGCCCAAGGAAGGACTTTAGAAATAGAAGATGCTTCTAGTCAGAAATTGCTCTTCTGATA	3102
Db	4923	GGGGTCCCTCGGAGGAGACTGATGTCGCCACGAGGTGTTGACCCGCAATTGAGTCATGCTCT	4982
Qy	3103	CTAGTTCAGGAGAAGGTGAA	3122
Db	4983	CTCGTCAAAAAGAAGGTGCA	5002

RESULT 15

US-10-006-285-400

Sequence 400, Application US/10006285

Publication No. US20030165854A1

GENERAL INFORMATION:

APPLICANT: Mary Jane Cunningham

APPLICANT: Matthew R. Kaser

TITLE OF INVENTION: MARKER GENES RESPONDING TO TREATMENT WITH TOXINS

FILE REFERENCE: PA-0039 US

CURRENT APPLICATION NUMBER: US/10/006,285

CURRENT FILING DATE: 2001-12-05

NUMBER OF SEQ ID NOS: 514

SOFTWARE: PERL Program

SEQ ID NO 400

LENGTH: 7805

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

OTHER INFORMATION: Incyte ID No. US20030165854A1 235885.5

US-10-006-285-400

Query Match	9.8%	Score 409.6	DB 15	Length 7805
Best Local Similarity	52.5%	Pred. No. 6.2e-98		
Matches 1240	Conservative 0	Mismatches 1024	Indels 96	Gaps 12
Qy	810	AGGCTTTACATCCATACACAGCTTGAGGACCTAAATTTTTTCGGGTTCTCGTGGTCAAA	869	
Db	2692	AGGCACCCTGCATGTATCAGTTGGAAGGGCTGAACCTGGCTACGCTTCTCTCTGGGCCCA	2751	
Qy	870	ACAGACGATGTTAATCCTTCTGATGAAATGGGACTAGGCAAGACAAATTCAAAGCATTTGC	939	
Db	2752	GGGCACCTGACACCAATCTAGCTGATGAGATGGGGCTAGGCAAGACCATACAAACCACATCGT	2811	
Qy	930	CCTTTTAGCTTTCACCTTTTTCAGGAGAACCTCA-----TTCCGCATTTGGTAATTGCTCC	983	
Db	2812	CTTCCCTCTACTCACTCTACAAGGAGGGGCCACACAAAGAGTCCCTTCTCGTGGTAGTGCCCC	2871	
Qy	984	TCCTATCGACTCTCGGTAACCTGGGAGAGAGAGTTTGCCACATGGGCCCCACAGATGACCT	1043	
Db	2872	ACTCTCTACCATCATTAACCTGGGAGGGGAGTTCAGATGTGGGCCACCCCAAAATTCATGT	2931	
Qy	1044	GGTTATGTATTTTGGCCTCGCAAGCTCGACGAGTTTATCAGAGAAACATGAGTTTTACTTT	1103	
Db	2932	GGTGACATACACGGGTGAAGGACGCCGGGCCATCTTCGTGAGATGAATTCCTCTT	2991	
Qy	1104	ATCGAAAGAT-CAAAAAAGATCAAGAAAAAGAAATCTCGACAAATTAAGTAGCGCAAGCA	1162	
Db	2992	TGAGGCAATGCCATCAAGGGGGCAAGAAAGCTTTTAAGATGAAGGGAGGCACAGG-	3050	
Qy	1163	AGCAAAAAAGAAATCAAGCTTTTGATGTCTCTCACATCGTATGAGATGATCAACCTAGATT	1222	

3051	Db	-----TGAAGTTCCATGTTTCTCCTGACATCGTATGAGCTGATCACCAATTGATC	3099
1223	Qy	CAGCAGTTCTAAAAACCAATTAAGCTGGGAGTGCATGATTTGTATGAAGGTCATCGACTGA	1282
3099	Db		
3099	Qy	AGGCAGCACTTGGTTCCATCCGCTGGCCCTGCTCTTGTGTAGATGAGGCCCATCGACTCA	3158
1283	Qy	AAATTAAGGANTCAAAAGCTGTTCTCTTTCAATPTGACACAGTANTCAAGTAAACCCACCGTATTC	1342
3159	Db	AGAACCAACCACTCCAAAGTTTTTTCAGGGTTTCTCAATGGTTTACAAGATAGATCATAAAGTTGC	3218
1343	Qy	TTCTGCAGAGAACACCACTTTCAGAAACAACTTGATGAACTTTTTCATGCTCATGCAATTTTC	1402
3219	Db	TGCTGACAGGAACCCCATTTGCAGAAATACTGGAGGAGCTCTTCCATCTCCTGAACCTTC	3278
1403	Qy	TTGATCGGGGAAAGTTTGGAAAGTTTGGAGGAGTTCAGGAGGAGTTCAAGATATTAAATC	1462
3279	Db	TCACCCACAGAGATTTAAACAATTCGAGGGCTTCCTGGAGGAGTTTGTGCATATATCCA	3338
1463	Qy	AAGAGAGCAGATCTCAAGGTTTGCAAAATGTTTGGCTCCACHTTTGTCTCAGAAGGGTAA	1522
3339	Db	AAGAGACCAAGATCAAGAAATCGCATGATTTGCTGGGGCCACACATGCTCGGAGACTCA	3398
1523	Qy	AAAAAGACGTTAATGAAGACATGCCCCCAAAAGAGAGCTCATTTTTCGTTGTTGATCTGA	1582
3399	Db	AGCGCATGTCTTTAAGAACATGCCAGCCAAAGACAGAGCTCATGTTTCGGGTGGAGCTAA	3458
1583	Qy	GCAGTCTGCAGAAAGAAATATTACAAAGCTATTTTTACCCTGAATTAATCAAGTATTGA---	1639
3459	Db	GCCCCATGCAGAGAAATACTACAAATACATCCTGACTCGAAATTTTGGAGCCCTGAAATT	3518
1640	Qy	CAAAAAAGGAGGTGCTCAAAATTTCCCTTAATAACATTATGATGGAATTAAGAAAAGTAT	1699
3519	Db	CACGAGGTGGTGGGAACCAAGGTGTCGCTGCTTAATATCATGATGGATCTTTAAGAAAGTGCT	3578
1700	Qy	GCTGCCATCTCTATATGCTAGAGGGTGTG-----AGCCAGTATTTCACGACGCA	1750
3579	Db	GCAACCATCCATACCTTTTCCGTGCTGCTATGAGTCCCCCAACTCCCCAGTGGGG	3638
1751	Qy	ATGAAGCTTTCAAACAACATTTTGGAGTCTGTGGAAAGCTGCACCTTCTAGATAAAATGA	1810
3639	Db	CTTATGAGGGTGGGGCACATTATTAAAGTCGCTGGGAAGCTCATGCTCTCCAGAAGATGC	3698
1811	Qy	TGCTCAAACTGAAGAGCAAGACACAGAGTCTCTATATACACACAGTTTCAGCATATGC	1870
3699	Db	TGGCAAGCTGAAGGACAGGACACCGAGTGTCTCATCTTCGCGAGATGACCAAAATGT	3758
1871	Qy	TGACTTACTTTGAAGACTACTGTACCATAAGAAATGGCAGTACGAGCGAAATGTATGGAA	1930
3759	Db	TAGACTTGCTTGAGGACTTCTTAGACTATGAAGGCTACAAGTATGAGCGCATCGATGGTG	3818
1931	Qy	AGTTGGCGAGCTGAGCGGCAAAATACGATAGATCGGTTCAATGCGCAAAAATTTCTAACA	1990
3819	Db	GTATACGGGTGCCCTTGAGGCAAGGAGCCATCGATCGGTTTAAATGCTCTCGGGGCCCAAC	3878
1991	Qy	AGTTTGTGTTTTTGTCTCCACAAGAGCTGGTGGCTTAGGAATAAATCTTGGCAACGGCTG	2050
3879	Db	AATCTGCTTCCTCTGTCCACCCGAGCTGGGGGCTTGGGCAATCAATCTGGCCACTGCTG	3938
2051	Qy	ATACAGTAAATCATTTATGACAGTGACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTA	2110
3939	Db	ACACTGTCATCTTTGATTCTGACTGGAAACCCCATAAATGACATCCAGGCCCTTTAGCC	3998
2111	Qy	GAGCTCATC-GACTTGGCCAAAACAAATAGAGTGATGATTTATAGGCTCATAAACCGAGGC	2169
3999	Db	GGGCTCATCGGATGGGCCAGGCCAACAAAGTGATGATTTACCGGTTTGTGATCGCGCG	4058
2170	Qy	ACCATTGAAGAAAGGATGATGAAATTTGACTAAAAAGAAATGGTTCTAGAGCATCTTGTT	2229
4059	Db	TCAGTGGAAAGCGGAATCACAAAGTGGCCAAAGAAAGATGATGCTGACACACCTTGGTT	4118
2230	Qy	GTTGG-----GAAACTCAAAACACAAAAACATTTAATCAGGAAGAGTTAGATGACATC	2280

Db 4119 GTGGGCTGGCTGGCTCCAGGCAGGCTCCATGTCCAGCAGGAGCTTGACGACATT 4178
 Qy 2281 ATCAGGTATGATCAAGAGGAGCTTTTCTAGTGAAGATGATGAAGCAGGAAAGTCTG-- 2338
 Db 4179 CTCAAATTTGGCATTGAAGAGCTATTCAAGGATGAAAACGAGGGGGAGAACAGAGGAG 4238
 Qy 2339 -----GAAAAATTCATTATGATGCGGCTATAGACAAATTCGTTGATCGTGATCTC 2391
 Db 4239 GACAGCAGTGTGATTCAATATGACAATGAGCCATCGCTGGCTGTTGGACCGGAACCG 4298
 Qy 2392 GTGAGGCAGAGGAGTCTCAGTGGATGATGAAGAGGAGAAATGAATCTTAAAGGCTTTC 2451
 Db 4299 GATCAACTGAGGA-----CACTGACGTGCAGAAACATGAATGATATCTCAGCTCCTTC 4352
 Qy 2452 AAGGTGCTAATTTTGAATATATAGATGAAAATGAGGCAGCAGCATTTAGAGGCACAGAGA 2511
 Db 4353 AAGGTGGCACAGTACGTCGTGCGGGAAGAGACAAGATTGAGGAAATTTAGCGGAGATC 4412
 Qy 2512 GTCGCTGCTGAAAGCAAATCTTCAGCAGGCAATTTCTGATAGAGCAAGTTATTGGGAAGAG 2571
 Db 4413 ATCAAGCAGGAGAGATGTGACCCCTGAC-----TACTGGGAGAAG 4454
 Qy 2572 TTGTTAAAGATAAATTTGAGCTGCACAGGCTGAGGAGCTTAATGCTCTTTGGAAGG 2631
 Db 4455 CTGCTGAGGCATCACTATGACAAACAGCAGGAGAACCTAGCCCGGAATCTAGGCAAGGC 4514
 Qy 2632 AAGAGAAGTCGACAGCAGTTGGTATCCATTGAAGAAGATGATCTTGCTGGTTTGGAGAT 2691
 Db 4515 AAGCGGTTCCCAAGCAAGTTAACTACAATGATGCTCTCAGGAAGACCAGCAACACGAG 4574
 Qy 2692 GTGAGCTCTGATGAGATGAAAGTTATGAAGCTCAGTCAACAGATGGTGAAGCAGCAGGA 2751
 Db 4575 TCAGAGTACTCGGTGGTTTCAGAGGAGGAGGATGAAGATTCGATGAACGCTCTGAAGGG 4634
 Qy 2752 CAAGGAGTTACAGCGGTTCGACGGCCGTACAGAGAAAGGTCGCGATAATTTGGAACCA 2811
 Db 4635 CGTAGACAGTCAAGAGGCGAGCTCCGAA-----TGAGAAAGATAAGCCACTG 4682
 Qy 2812 ACTCGTTGATGGAAGTGAGGGGAGATCTTTACAGACTACTGGGTTTCAACCAGAGTCAA 2871
 Db 4683 CCTCCACTGCTGGCCCGAGTCGGGGGCAACATTGAGGTGCTTGGGCTTCAACACCCGTCAG 4742
 Qy 2872 AGGGCCATTTTGTACAGACTTTGATGAGGTATGG-----AGCTGGCAATTTTGTAT 2922
 Db 4743 CGGAAGGCTTTCTCAATGCTGTGATGCGCTGGGGATGCCACCACAGGATGCCCTTCACC 4802
 Qy 2923 TGAAGGAGTTTGTCTCGCTTAAAGCAGAACCTTTTGAAGAAATAAATGAATATGGA 2982
 Db 4803 ACACAGTGGCTGCTGCGGGACCTGAGGGGCAAGACTGAGAGGAGTTTAAAGGCCATATGTG 4862
 Qy 2983 ATACTCTTTGAAGCAGATTTGCTGAAGAAATAGACGAGATCTCCAACTTTTCAGAT 3042
 Db 4863 TCTTTGTTTCATGCGCCATCTCTGTGAGCCTGGGGCAGACGGCTCTGAACCTTTTGGCCGAT 4922
 Qy 3043 GGTGTGCCCAAGGAGGACTTTAGAATAGAAGATGTTCTAGTCAGAATTTGCTTCTTGATA 3102
 Db 4923 GGGGTCCCTCGGAGGAGACTGAGTCGCCAGCAGGTGTTGACCCGATTTGGAGTCATGTCT 4982
 Qy 3103 CTAGTTACGAGAGGTTGAA 3122
 Db 4983 CTCGTCAAAAAGAGGTGCA 5002

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 14:48:31 ; Search time 9894 Seconds
(without alignments)
12607.068 Million cell updates/sec

Title: US-10-049-137-1
Perfect score: 4177
Sequence: 1 atgagtagtttggtggagag.....caactggtaaatcaagattc 4177

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: em_gss_hum:*
 - 18: em_gss_inv:*
 - 19: em_gss_pln:*
 - 20: em_gss_vrt:*
 - 21: em_gss_fun:*
 - 22: em_gss_man:*
 - 23: em_gss_mus:*
 - 24: em_gss_pro:*
 - 25: em_gss_rpd:*
 - 26: em_gss_png:*
 - 27: em_gss_vrl:*
 - 28: gb_gss1:*
 - 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	910.2	21.8	939	14	CF651673 08-L02052
2	612.8	14.7	639	9	AU237917 AU237917
3	528	12.6	528	9	AV551529 AV551529
4	415.4	9.9	4226	11	BC059082 BC059082 Mus muscu

5	408.2	9.8	3137	11	AK034549	AK034549 Mus muscu
6	398.8	9.5	610	10	BE204297	BE204297 EST396973
7	387.6	9.3	745	14	CB635984	CB635984 OSI1B16P
c	385	9.2	563	9	A1998628	A1998628 701546371
8	373	8.9	597	10	BE204298	BE204298 EST396974
10	354.6	8.5	589	13	B0139596	B0139596 NF021H12P
11	351.4	8.4	583	10	BE205450	BE205450 EST398136
12	336	8.0	778	14	CA240279	CA240279 SCSBFL406
13	308	7.4	348	13	BQ834228	BQ834228 ALEST0190
c	300	7.2	503	28	B29916	B29916 T1D20TRB T
15	298.2	7.1	673	14	CD812910	CD812910 BN10.022P
16	296.2	7.1	4209	11	AK049686	AK049686 Mus muscu
17	287.8	6.9	557	9	AU287846	AU287846 AU287846
18	285.6	6.8	2740	11	AK039580	AK039580 Mus muscu
19	280.2	6.7	492	9	AU225820	AU225820 AU225820
20	278	6.7	843	14	CB667080	CB667080 OSJNED14E
21	277.2	6.6	655	12	B1928046	B1928046 EST547935
22	276.8	6.6	673	12	BM406422	BM406422 EST580737
23	276	6.6	653	13	BU000377	BU000377 QGG24K16
24	272.4	6.5	680	13	BQ518530	BQ518530 EST625945
25	267.4	6.4	583	9	AV835455	AV835455 AV835455
26	265.6	6.4	656	14	CB079334	CB079334 hp6907.b
27	257.4	6.2	810	13	CA077816	CA077816 SCRFAM102
28	253.4	6.1	1811	11	AK015218	AK015218 Mus muscu
29	253.2	6.1	2877	11	AK030430	AK030430 Mus muscu
30	251.6	6.0	3161	11	AK040495	AK040495 Mus muscu
c	251.4	6.0	830	13	BX748807	BX748807 BX748807
32	242.8	5.8	3607	11	AK018129	AK018129 Mus muscu
33	235.2	5.6	752	14	CB525972	CB525972 UI-M-FY0-
34	233.8	5.6	502	12	BJ186506	BJ186506 BJ186506
35	232.4	5.6	532	14	CA170938	CA170938 SCRUSB106
36	232.4	5.6	874	13	BU194813	BU194813 AGENCOURT
37	231	5.5	935	9	AL519189	AL519189 AL519189
38	228.4	5.5	813	14	CA324135	CA324135 UI-M-FY0-
39	227.8	5.5	767	14	CD349796	CD349796 UI-M-FY0-
40	226.2	5.4	675	14	CA500458	CA500458 WHE4019.H
41	225.8	5.4	775	14	CB525974	CB525974 UI-M-FY0-
c	223.8	5.4	818	14	CF784972	CF784972 AGENCOURT
43	223.6	5.4	516	12	B1321298	B1321298 saf50b11
44	223.4	5.3	766	14	CB247361	CB247361 UI-M-FY0-
45	221.4	5.3	2168	11	AK052092	AK052092 Mus muscu

ALIGNMENTS

RESULT 1	CF651673	939 bp	mRNA	linear	EST 06-NOV-2003
LOCUS	08-L020524-066-003-002-SF6P	MP1Z-ADIS-066	Arabidopsis thaliana	cDNA	
DEFINITION	clone MP1Zp20010023Q 5-PRIME, mRNA sequence.				
ACCESSION	CF651673				
VERSION	CF651673.1	GI:37427433			
KEYWORDS	EST				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Schmid,K.J., Soerensen,T.R., Stracke,R., Torjek,O., Altmann,T., Mitchell-Olds,T. and Weissshaar,B.				
AUTHORS	Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in Arabidopsis thaliana				
TITLE	Genome Res. 13 (6), 1250-1257 (2003)				
JOURNAL	22683290				
MEDLINE	12799357				
PUBMED	Contact: Weissshaar B				
COMMENT	ADIS DNA core facility at MP1Z Max-Planck-Institute for Plant Breeding Research Carl-von-Linne Weg 10, 50829 Koeln, Germany Fax: 00492215062851 Email: weissshaar@mpiz-koeln.mpg.de				

Insert Length: 939 Std Error: 0.00

Plate: 3 row: 0 column: 02

Seq primer: SP6P:

FEATURES
source

Location/Qualifiers

1. 939
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/cultivar="Wassilewskija-0 (Ws-0)"
/db_xref="GABI:938264"
/db_xref="taxon:3702"
/clone="MPI2p20010023Q"
/tissue_type="root"
/lab_host="E. coli TOP10"
/clone_lib="WP12-ADIS-066"
/note="vector: pCMVSPORT6; Site 1: Sall; Site 2: NotI;
cDNA library from Arabidopsis thaliana, accession
Wassilewskija-0; roots from three weeks old plants grown
on MS-plates at 26M-OC with 16 hours light/day; library
was made at the Max-Planck-Institute for Plant Breeding
Research, Cologne, Germany; cloning sites Sall-NotI,
primer sites and orientation:
SP6-Sall-CCACGCGCCG-5prime-cDNA-polyA-CC-NotI-T7; GATEWAY
compatible; Note: Sequencing granted in the context of the
GABI Arabidopsis Verbund I: Genetic Diversity,
'Establishment of high-efficiency SNP-based mapping tools
and development of methods for genome-wide mutation
detection' PI: Bernd Weisshaar Sequence submission managed
by RZPD/GABI-Primary database: http://gabi.rzpd.de This
clone is available from RZPD; contact RZPD (clone@rzpd.de)
for further information."

ORIGIN

Query Match 21.8%; Score 910.2; DB 14; Length 939;
Best Local Similarity 99.5%; Pred. No. 7.5e-195;
Matches 923; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 1 ATGAGTAGTTGGTGGAGAGCGCTTCGCATACGATCTGATAGGAACACGAGTTTAAACCTA 60
Db 13 ATGAGTAGTTGGTGGAGAGCGCTTCGCATACGATCTGATAGGAACACGAGTTTAAACCTA 72
Qy 61 GATGATTCGTGATGATGACGATTCGTTCTTCTTAAAAAGATCGAATTCGAGTTCGAG 120
Db 73 GATGATTCGTGATGATGACGATTCGTTCTTCTTAAAAAGATCGAATTCGAGTTCGAG 132
Qy 121 GCTATTGTCGAACTGATGCGAAGAAATGCATGCTCAGGCTGCTGGGGAAGTACTAAT 180
Db 133 GCTATTGTCGAACTGATGCGAAGAAATGCATGCTCAGGCTGCTGGGGAAGTACTAAT 192
Qy 181 CTGTGAAGCTGCAATACATGACATTCATGCGGTTCCATGCTAAATGCTTAGTTCCACCTCTT 240
Db 193 CTGTGAAGCTGCAATACATGACATTCATGCGGTTCCATGCTAAATGCTTAGTTCCACCTCTT 252
Qy 241 AAAGATGCTTCGCTGGGAAATTTGAGATGCGCTCGAATGTTGTTAGTCTCTTTAAACGAGATA 300
Db 253 AAAGATGCTTCGCTGGGAAATTTGAGATGCGCTCGAATGTTGTTAGTCTCTTTAAACGAGATA 312
Qy 301 GATAGATATTGGATTGGATGGCTTCCTACAAATCTAGTGAACAGGTTCTCCGAT 360
Db 313 GATAGATATTGGATTGGATGGCTTCCTACAAATCTAGTGAACAGGTTCTCCGAT 372
Qy 361 GCGGAACCGAAGCCAAATTTTGTGAACAGATATCTCGTGAAGTGGGAAGGATTTATCATAC 420
Db 373 GCGGAACCGAAGCCAAATTTTGTGAACAGATATCTCGTGAAGTGGGAAGGATTTATCATAC 432
Qy 421 CTTCACTGCTCTTGGGTGCTCGAGAGAGGTTCCAGAAAGGCTTTAATAGTCAAAATCATCGT 480
Db 433 CTTCACTGCTCTTGGGTGCTCGAGAGAGGTTCCAGAAAGGCTTTAATAGTCAAAATCATCGT 492
Qy 481 TTAARAAACGAGTGAACAATTTTCCCGTCAAAATGGAGTCAATTAACAGCGAGAGAT 540
Db 493 TTAARAAACGAGTGAACAATTTTCCCGTCAAAATGGAGTCAATTAACAGCGAGAGAT 552
Qy 541 GATTTTGTGCCATACGTCCTGAGTGGGACCACTGTGTGATCGGATTTCTTGCCTGCAGAGAG 600

Db 553 GATTTTGTGCCATACGTCCTGAGTGGACCACTGTTGATCGGATTTCTTGCCTGCAGAGAG 612
Qy 601 GAGATGGGGAGCTGCAATATCTTGTCAAAATATAAAGAGCTATCCTATGATGAATGTAT 660
Db 613 GAAGATGGGGAGCTGCAATATCTTGTCAAAATATAAAGAGCTATCCTATGATGAATGTAT 672
Qy 661 TGGGAGTGCAGAATCAGACATCTCAACCTTCCAGAAATTCAGAAAGTTCAAGGATGTA 720
Db 673 TGGGAGTGCAGAATCAGACATCTCAACCTTCCAGAAATTCAGAAAGTTCAAGGATGTA 732
Qy 721 AATTCTAGAACTCCGAGAAGTAAAGATGTTGACCAATAAAAGAAATCCAGAGACTTTCAA 780
Db 733 AATTCTAGAACTCCGAGAAGTAAAGATGTTGACCAATAAAAGAAATCCAGAGACTTTCAA 792
Qy 781 CAGTTGATCATACTCTGTAATTCCTCAAGGGTTGTACATCCATACACGCTTGAGGGA 840
Db 793 CAGTTGATCATACTCTGTAATTCCTCAAGGGTTGTACATCCATACACGCTTGAGGGA 852
Qy 841 CTTAAATTTTGGGGTCTCGTGGTCAAAACAGACGCGATGTAATCCTTGTGATGAAATG 900
Db 853 CTTAAATTTTGGGGTCTCGTGGTCAAAACAGACGCGATGTAATCCTTGTGATGAAATG 911
Qy 901 GGACTAGGCAAGCAATTCAAAACATG 928
Db 912 GGACTAGGCAAGCAATTCAAAACATG 939

RESULT 2
AU237917
LOCUS
DEFINITION AU237917 RAFL16 Arabidopsis thaliana cDNA clone linear EST 01-APR-2002
mRNA sequence.
ACCESSION AU237917
VERSION AU237917.1 GI:19877086
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 639)
AUTHORS Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
Akiyama,K., Enju,A., Oono,Y., Sakurai,T., Carninci,P., Kawai,J.,
Itoh,M., Iehi,Y., Arakawa,T., Shibata,K., Shinagawa,A.,
Muranatsu,M., Hayashizaki,Y. and Shinozaki,K.
Large scale analysis of Arabidopsis full-length cDNA
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msek@rtc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et
al., submitted for publication) digested with BamHI and Sall. This
clone is in a modified pBluescript vector. Please visit our web
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.

FEATURES
source
1. 639
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAFL16-66-N20"
/lab_host="DH10B"
/clone_lib="RAFL16"
/note="Site_1: BamHI; Site_2: Sall; dark-grown"

ORIGIN

Query Match 14.7%; Score 612.8; DB 9; Length 639;
Best Local Similarity 99.4%; Pred. No. 1.3e-127;
Matches 625; Conservative 0; Mismatches 37; Indels 1; Gaps 1;

QY 1254 CATGATTGTTGATGAAGGTCATCGACTCGAAATAAAGGATTCAAAGCTGTTCTCTTCATT 1313
DB 1 CTTGATTGTTGATGAAGGTCATCGACTCGAAATAAAGGATTCAAAGCTGTTCTCTTCATT 60

QY 1314 GACACAGTATTCAGTAACACCGTATCTTCTGACAGGAACACCACTTCAGAACAACTT 1373
DB 61 GACACAGTATTCAGTAACACCGTATCTTCTGACAGGAACACCACTTCAGAACAACTT 120

QY 1374 GGATGAACCTTTTCATGCTCATGCTATCTTCTGATGCGGGAAGCTTGGAGTTTGAGGA 1433
DB 121 GGATGAACCTTTTCATGCTCATGCTATCTTCTGATGCGGGAAGCTTGGAGTTTGAGGA 180

QY 1434 GTTCCAGGAGGAGTTCAAAGATATTAATCAAGAGGAGCAGATCTCAAGGTTGCACAAAT 1493
DB 181 GTTCCAGGAGGAGTTCAAAGATATTAATCAAGAGGAGCAGATCTCAAGGTTGCACAAAT 240

QY 1494 GTTGGCTCCACATTTGCTCAGAAGGTAAAGAGCGTAATGAAAGACATGCCCCCAA 1553
DB 241 GTTGGCTCCACATTTGCTCAGAAGGTAAAGAGCGTAATGAAAGACATGCCCCCAA 300

QY 1554 AAAGGAGCTCATTTTGGCTGCTGATCTGAGCAGTCTGCAGAGAAATATTAACAAGCTAT 1613
DB 301 AAAGGAGCTCATTTTGGCTGCTGATCTGAGCAGTCTGCAGAGAAATATTAACAAGCTAT 360

QY 1614 TTTTACCGTAAATATCAAGTATTTGACAAAAGGAGGAGTGTCAAATTTCCCTTAATAA 1673
DB 361 TTTTACCGTAAATATCAAGTATTTGACAAAAGGAGGAGTGTCAAATTTCCCTTAATAA 420

QY 1674 CATTATGATGAATTAACGAAAGTATGCTGCCATCTTATATGCTAGAGGGTGTGAGCC 1733
DB 421 CATTATGATGAATTAACGAAAGTATGCTGCCATCTTATATGCTAGAGGGTGTGAGCC 480

QY 1734 AGTTATTCAGCAGCAAAATGAAGCTTCAACAACTTTTGGAGTCTGTGGAAAGCTGCA 1793
DB 481 AGTTATTCAGCAGCAAAATGAAGCTTCAACAACTTTTGGAGTCTGTGGAAAGCTGCA 540

QY 1794 ACTTCTAGATAAAATGA-TGGTCAAATGAAGACGACACAGAGTCTTAATATACA 1852
DB 541 ACTTCTAGATAAAATGA-TGGTCAAATGAAGACGACACAGAGTCTTAATATACA 600

QY 1853 CACAGTTTCAGCATATGCTGGACTTACTT 1881
DB 601 CACAGTTTCAGCATATGCTGGACTTACTT 629

RESULT 3
AV551529
LOCUS Arabidopsis thaliana (thale cress)
DEFINITION Arabidopsis thaliana
VERSION AV551529.1 GI:8722942
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana

REFERENCE
AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
JOURNAL DNA Res. 7 (3), 175-180 (2000)
MEDLINE 20363093
PUBMED 10907847
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

Location/Qualifiers
1..528
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RZ127h1R"
/tissue_type="roots"
/clone_lib="Arabidopsis thaliana roots Columbia"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 12.6%; Score 528; DB 9; Length 528;
Best Local Similarity 100.0%; Pred. No. 1.9e-108;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2759 TTCAGACGGGTCGACGGCGGTACAGAGAAAGGGTCGCGATTAATTTGGAAACCACTCCGT 2818
DB 1 TTCAGACGGGTCGACGGCGGTACAGAGAAAGGGTCGCGATTAATTTGGAAACCACTCCGT 60

QY 2819 TCATGGAAGGTCAGGGGAGATCTTTCAGAGTACTGGGTTTCAACCAAGTCAAAAGGCCA 2878
DB 61 TCATGGAAGGTCAGGGGAGATCTTTCAGAGTACTGGGTTTCAACCAAGTCAAAAGGCCA 120

QY 2879 TTTTGTGACAGACTTTGATGAGGTATGGAGCTGGCAATTTGATTGGAAGGAGTTTGTTC 2938
DB 121 TTTTGTGACAGACTTTGATGAGGTATGGAGCTGGCAATTTGATTGGAAGGAGTTTGTTC 180

QY 2939 CTCGCTTAAAGCAGAGACCTTTTGAAGAAATAAATGAATATGGAATCTCTTCTGAAGC 2998
DB 181 CTCGCTTAAAGCAGAGACCTTTTGAAGAAATAAATGAATATGGAATCTCTTCTGAAGC 240

QY 2999 ACATTGCTGAAGAAATAGACGAGATTTCTCCAACTTTTCAGATGCTGTCGCCAAGGAG 3058
DB 241 ACATTGCTGAAGAAATAGACGAGATTTCTCCAACTTTTCAGATGCTGTCGCCAAGGAG 300

QY 3059 GACTTAGAATAGAAGATGTTCTAGTCAGAAATTTGCTCTGATATCTAGTTTCAGGAGAAG 3118
DB 301 GACTTAGAATAGAAGATGTTCTAGTCAGAAATTTGCTCTGATATCTAGTTTCAGGAGAAG 360

QY 3119 TGAATTTGTAGAAGATCATCCAGGAAACCTGTTTTCCCTCTCGATCTTGTGAAGAT 3178
DB 361 TGAATTTGTAGAAGATCATCCAGGAAACCTGTTTTCCCTCTCGATCTTGTGAAGAT 420

QY 3179 TCCCGGACTGAGAGTGGAAAAATTTGGAAGGAGGACATCACAAGATAATGATACGTG 3238
DB 421 TCCCGGACTGAGAGTGGAAAAATTTGGAAGGAGGACATCACAAGATAATGATACGTG 480

QY 3239 CTGTTTTAAAGCATGGGTACGACGGTGGCAAGCTATTGTTGATGACA 3286
DB 481 CTGTTTTAAAGCATGGGTACGACGGTGGCAAGCTATTGTTGATGACA 528

RESULT 4
BC059082
LOCUS Mus musculus cDNA clone IMAGE:6827288, containing frame-shift errors.
DEFINITION BC059082
VERSION BC059082.1 GI:37590750
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

[illegible]

Db 1242 GGAACATGCGTCAGGAGCCATTGATGCTTCAATGCACCGGTGCTCAACAGTTCTGCT 1301
Qy 2000 TTTTGTCTCTCCAAAGAGCTGGTGTAGGAATAAATCTTGCAACGGCTGATACAGTAA 2059
Db 1302 TCTTGTCTTCCACTCGAGCTGGGGGCTTGGGATCAATCTGGCCAATCTGGGACACAGTTA 1361
Qy 2060 TCATTATGACAGTGAGTGAATCCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
Db 1362 TTATATATGACTCTGACTGGAACCCCATTAATGA-CATCAGGCTTTAGCAGAGCCACC 1420
Qy 2120 GACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCATAAACCGAGGCCACCATTTGAAG 2179
Db 1421 GTATTGGCCAAATAAGAGTGTATCTACCGTTTGTGACCGTGCATCAGTGGAG 1480
Qy 2180 AAAGGATGATGCAATTGACTAAAGAAATGTTCTTAGAGCATCTTGTGTTGG----- 2234
Db 1481 AGCGCATACAGAGTGGCAAGAAAGATGATGCTGCACATTTAGTGTACGCGCTG 1540
Qy 2235 ----GAAACTCAAAACACAAAACATTAATCAGGAAGTTAGATCATCATCAGTATG 2290
Db 1541 GCTGGGCTTCAAGACAGGCTCAATGTCTAAACAGGAGCTTGATGATATCCTCAAAATTG 1600
Qy 2291 GATCAAGGAGCTTTTGTGCTAGTGAAG-----ATGATGAAGCAG 2329
Db 1601 GCACTGAGGAGCTATTCAAGGATGAAGCCAGGATGGAGGAGACACAGGAGGAG 1660
Qy 2330 GAAAGTCTGAAAAATTCATTATGATGCGCTATAGACAAATTCGTTGATCGTGAATC 2389
Db 1661 AAGACAGCAGTGTATCCACTATGACGATAAGGCCATTGAACGACTGCTGGATCGAAACC 1720
Qy 2390 TCGTGAGGCGAGGAGTCTCAGTGATGATGAAGAGGAGATGATCTTTAAAGCTT 2449
Db 1721 AGATGAGACTGAAGACACAGAATTG-----CAGGGCATGAATGAATATTGAGCTCAT 1774
Qy 2450 TCAAGTGGCTTAATTTTGAATATATAGTAAAAATCAGGAGCAGCATTAGAGGCACAGA 2509
Db 1775 TCAAGTGGCTCAGTATGTGTGACGGAAGAGAGATGG-----GGGAGGAGAGG 1825
Qy 2510 GATCGCTGCTGAAACAAATCTTACGAGGCAATCTGATAGCAAGTTATTGGGAAG 2569
Db 1826 AGGTAGAACGGAAATCATAAACAGAGAAAGTGTGGATCTGAC---TACTGGGAGA 1882
Qy 2570 AGTTGTTAAAGATAATTTGAGCTCCAGCTGAGGCTTATGCTCTTGGAAAAA 2629
Db 1883 AATTGCTGGCGACCAATTATGACGACGACGAAGAATCTAGCCGAAATCTGGGCAAG 1942
Qy 2630 GGAAGAGAAGTCCCAAGCAGTTGGTATCCATTGAAG---AAGATGATCTTGTGTTGG 2686
Db 1943 GAAAAAGATCCGTAACAGGTCACACTACATGATGCTCACAGGAGGCCGATGGC 2002
Qy 2687 AGATGTGAGCTCTGATGGAGATGAAAGTTATGAAGCTGAGTCAACAGATGGTGAAGCAG 2746
Db 2003 AGGACGACAGTCCGACCAACAGTCCGATTTACTCAGTGGCTCAGAGGAAGGTGATGAAG 2062
Qy 2747 CAGGACAAAGGATTCAGCGGTGCGAGCGCCGTACAGAGAAGGG-----TCGCG 2797
Db 2063 ACTTTGACCAACGGTCAGAGCTCCCGCGAGGCCAGTCCGACAGGCGCTCGGAATGATA 2122
Qy 2798 ATAAATTTGAACCAACTCCGTTTCATGGAAGTTCAGGGGAGATCTTTTCAGAGTACTGGGT 2857
Db 2123 AGATGAAGCATTAATCTCTCTGTTGGCCGTTGTTGGGATATTTGAATGACTTGGTT 2182
Qy 2858 TCAACAGAGTCAAAAGGCCATTTTGTACAGACTTTGATGAGGTATGG-----AG 2908
Db 2183 TTAATGCTGTCAGCGAAAGACATTTCTTAATGCAATTTATGATATGCAATGCCACCTC 2242
Qy 2909 CTGGCAATTTGATGGAAGGATTTGTTCTCGCTTAAAGCAGAGACCTTTGAAGAAA 2968
Db 2243 AGGATGCTTTTACCACTCAGTGGCTTGTGAGAGATCTTCAGGCAAGTCAGAGAAAGAT 2302
Qy 2969 TAAATGAATATGAATACTCTTCTGAAGCAGACATTTCTGAAGAAATAGACGAGATTTCTC 3028
Db 2303 TTAAGGCTTATGTGTCACCTCTTCTATCGGACATTTGTGTGAGCCTGGGCGAGTGGGCTG 2362

Qy 3029 CAACCTTTTCAGATGTTGCCCAAGGAAGGACTTAGAATAGAAAGATGTTCTAGTCAGAA 3088
Db 2363 AGACCTTTTCTGATGGCGTCCACGAGAGGCGCTGTCTCGTCAACAGTTTCTCACTAGGA 2422
Qy 3089 TTGCTCTTCTGATACTAGTTTCAGGAGAGGTCGAAATTTGTAGAAGAT 3135
Db 2423 TCGGTGTCATGCTCTGATTCGAAAGAGGTTTCAGGAGTTTGAACAT 2469

RESULT 5
AK034549
LOCUS
DEFINITION
Mus musculus 12 days embryo embryonic body between diaphragm region
and neck cDNA, RIKEN full-length enriched library, clone:9430004K15
product:CHROMODOMAIN HELICASE-DNA-BINDING PROTEIN 4 (CHD-4) (MI-2
AUTOANTIGEN 218 KDA PROTEIN) (MI2-BETA) homolog (Homo sapiens),
full insert sequence.
AK034549
VERSION AK034549.1 GI:26330020
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, S., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
PUBMED 11076861
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
PUBMED 12031111
REFERENCE 6
AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsumi, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,

TITLE JOURNAL	Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akashira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-Ku, Yokohama, Kanagawa 230-0445, Japan. (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	Db	1474	ATACTATCTTGCTGATGAGATGGGCTTGGGAAACCTGTGCGAGACAGTCTTCTCT 1533
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/.	Qy	938	CTTCACTTTTTCAGGAGAACCTCAAT-----CCGCAATTTGGTAATTTGCTCTATCGA 991
FEATURES source	Location/Qualifiers 1. .3137 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:9430004K15" /db_xref="MGI:2398754" /db_xref="taxon:10090" /clone="9430004K15" /tissue_type="embryonic body between diaphragm region and neck" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="12 days embryo" 1. .3137 /note="unnamed protein product; CHROMODOMAIN HELICASE-DNA-BINDING PROTEIN 4 (CHD-4) (MI-2 AUTOANTIGEN 218 KDA PROTEIN) (MI2-BETA) homolog (Homo sapiens) (SWISSPROT Q14839, evidence: PASTV, 99.8%ID, 54.6%length, match=3135) putative"	Db	1534	ACTCCCTCTATAGAGGGTCACTCAAAGGCCCTTCTTAGTGAGCGCTCTCTGTCCA 1593
CDS		Qy	992	CTCTGCGTAACCTGGGAGAGAGATTGGCCATATGGCCCCCAGATGAACGTGGTTATGT 1051
		Db	1594	CCATCATCACTGGGAGAGAGATTGAGATGTGGCTCTCTGATATGATGTGGTAACCT 1653
		Qy	1052	ATTTGGGACTCGGCAGCTCGAGAGTTATCAGAGAACATGAGTTTACTTATCAAAAG 1111
		Db	1654	ATGTGGTGACAAAGGACAGCCGTGCATCATCCGGGAAATGAGTTCTCTTTGAAGACA 1713
		Qy	1112	ATCAAAABRAGATCAAGAAAGAAATCTGGACAAATAAGTAGCGGAAACGACAAAGAA 1171
		Db	1714	ATGCCATTCGTGGTGGCAAGAGCCCTC-----GAGGATGAAGAAGAGCAT 1761
		Qy	1172	GAATCAAGTTTGATGTCTCTCCTCACATCGTATGAGATGATCAACCTAGATTGACGAGTTC 1231
		Db	1762	CGGTGAATTCATGTCTCTGCTGACATCCTATGAGTTGATCACCATTGATATGGCCATCT 1821
		Qy	1232	TAAACCAATTAAGTGGGAGTGCATGATTGTTGATGAAGGTCACTGCAGTGAAGAAATAGG 1291
		Db	1822	TGGGTTCTATTGACTGGGCTGCTCAITTTGTGGATGAAGCCCATCGGCTGAAGAACCAAC 1881
		Qy	1292	ATTCAAGCTGTCTCTCTTCAATTGACACAGTATTCAAGTAACCAACCGTATTCTTCTCACAG 1351
		Db	1882	AGTCTAAGTTCTTTCAGTTTGAATGGTTACTCTCTCCAGCACAAGCTGTGTGCTAATCTG 1941
		Qy	1352	GAAACCACTTCAGAAACAACTTTGGATGAACCTTTTCATGTCTCATGTCAATTTCTTGTATGCGG 1411
		Db	1942	GAACTCGTTACAGAACAACTTAGAGGAATGTTTTCATCTGTCTCAACTTTCTCACCCCTG 2001
		Qy	1412	GGAAGTTTGAAGTTTGGAGAGTTCAGAGAGAGTTTCAAGATATTAATCAAGAGAGC 1471
		Db	2002	AGAGGTTTCCAACTATTAGAAGGCTTCTTGAGGAGTTTTCAGACATTTGCCAAGAGGAGC 2061
		Qy	1472	AGATCTCAAGTTTGCAAAATTTGGCTCCACATTTGTCTCAGAGGGTAAAGAAAGAGC 1531
		Db	2062	AGATTAAATACTCAACGACATGCTGGGCTCATATGTTGCGCGCTCAAGCTGAGC 2121
		Qy	1532	TAAATAAGACATGCCCCCAAAAGAGAGCTCAATTTTGGTGTGTATCTGAGCAGTCTGC 1591
		Db	2122	TCTTCAAGAATATGCCATCCAACAGACAGCTGATTTGTCGCGTGGAGTTGAGTCTTATGC 2181
		Qy	1592	AGAAAGATATTCAAGAGTATTTTACCGTAAATTTATCAAGTATTGA---CAAAAAGG 1648
		Db	2182	AGAAGAAGTACTACAAAGTATATTCTCACGCGGAATTTTGAAGCATTAAATGCTCGAGGTG 2241
		Qy	1649	GAGGTGCTCAAAATTTCCCTTAATAACATTTATGATGAAATTTACGAAAGTATGCTGCATC 1708
		Db	2242	GTGGCAACAGGTTTCTGCTGAACTGGTGTATGATCTTAAGAATATGCTGCAACACC 2301
		Qy	1709	CTTA-----TATGCTAGAGGGTGTGAGCCAGTTATTACGACGCGAAATGAAGCTT 1759
		Db	2302	CTTATCTTCTCCCTGTGGAGCAATGGAAGCCCTTAAGATGCCTAATGGTATGTATGATG 2361
		Qy	1760	TCAACAACCTTTTGGAGTCTTGTGGAAAGCTGCACTCTTAGATAAATATGATGGTCAAC 1819
		Db	2362	GCAGTGCCTTAATCAGAGCATCTGGGAAGTTTGTGCTGCGAAGAGTCTTAAAGAAC 2421
		Qy	1820	TGAAGAGCAAGGACACAGAGTCTTAATATACACAGATTTTCAAGCATATCTGCACTTAC 1879
		Db	2422	TGAAGAGGAGGAGCATCTGTGCTCATCTTCTCCAGATGACCAAGATGTGGACTTGC 2481
		Qy	1880	TTGAAGACTACTGTACCCATTAAGAAATGGCAGTACGAGCGAAATGATGGAAGAGTTGGCG 1939
		Db	2482	TAGAGGATTTTGTAGAACAATGAAGTTATAAATATGAACGTAATGATGTTGGAATCACTG 2541
		Qy	1940	GAGCTGAGCGGCAAAATACGATAGATGGTTTCAATGCCAAAAATTTCTAACAAGTTTGT 1999
		Db	2542	GGAAACATGCGTCAAGGAGGCAATTTGATCGCTTCAATGCAACCGGGTGTCTCAACAGTTCTGCT 2601

ORIGIN

Query Match	9.8%; Score 408.2; DB 11; Length 3137;
Best Local Similarity	56.7%; Pred. No. 3.1e-81;
Matches	874; Conservative 0; Mismatches 628; Indels 39; Gaps 5;
Qy	818 TACATCCATACAGCTGAGGACTTAATTTTTTTCGGTCTCTCGTGGTCAAAACAGACGC 877
Db	1414 TGCATCCCTACCAATGAAGGCTTAATCTGGCTCCGCTTCTCTGGCTCAGGGAACCG 1473
Qy	878 ATGTAATCTTGTCTGATGAATGGGACTTAGGCAAGACAATTCAAAGCATTTGCCCTTTT 937

QY 2000 TTTTGCTCTCCACAGAGCTGGTGGCTTAGGAATAATCTTGCAACGGCTGATACAGTAA 2059
 Db 2602 TCTTGCTTTCACTCGAGCTGGGGGCTTGGGATCAATCTGGCCACTGCGGACACAGTTA 2661
 QY 2060 TCATTATGACAGTCACTGGAATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATC 2119
 Db 2662 TTATATATGACTGTGCTGAGACCCCAATATGATCATCCAGGCTTTAGCAGAGCCACC 2721
 QY 2120 GACTTGGCCAAACAAATAGAGTGATGATTTATAGGCTCATAAACCGAGGCCACATTTGAAG 2179
 Db 2722 GTATTGGGCAAAATAGAAAGTGATGATCTACCGTTTGTGACCCGCTGCATCAGTGGAGG 2781
 QY 2180 AAAGGATGATGCAATGACTTAAAGAAAGAAATGTTCTAGAGCATCTTGTTGG----- 2234
 Db 2782 AGCGCATCACGAGGTGGCAAGAAAGAAAGATGATGCTGACACATTTAGTGGTACGCGCTG 2841
 QY 2235 ----GAAACTCAAAACACAAACAAATTAATCAGGAAGAGTTAGATGATCATCATCAGGTATG 2290
 Db 2842 GCTGGGCTCAAGACAGGCTCAATGCTTAAACAGAGGCTTGATGATATCTCAAAATTTG 2901
 QY 2291 GATCAAGAGGAGCTTTTGTGCTAGTGAAGATGATGAAGCAGGA 2331
 Db 2902 GCATGAGGAGCTATTCAAGGATGAAGCCAGGATGGAGGA 2942

RESULT 6
 BE204297
 LOCUS
 DEFINITION BE204297 610 bp mRNA linear EST 05-SBP-2000
 EST396973 KVO Medicago truncatula cDNA clone PKV0-15G16, mRNA
 sequence.

ACCESSION BE204297
 VERSION BE204297.1 GI:8747582
 KEYWORDS EST.
 SOURCE Medicago truncatula (barrel medic)
 ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.

REFERENCE 1 (bases 1 to 610)
 VandenBosch, K., Endre, G., Hur, J., Moore, J., Beremand, P., Ellis, L.,
 Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.B. and
 Fraser, C.M.
 Unpublished (1999)
 ESTs from uninoculated seedling roots of Medicago truncatula

COMMENT
 Contact: VandenBosch K
 Department of Plant Biology
 University of Minnesota
 220 BioSci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
 Tel: 612 624 2755
 Fax: 612 625 1738
 Email: kvanden@cba.umn.edu

Texas A&M University name: T264069e
 TIGR sequence name: MTGAR44TK
 More information is available at:
<http://chrystie.tamu.edu/medicago>
 Seq primer: Sknmd (CTA gaa cta gta gta CC).

FEATURES
 source
 1..610
 /organism="Medicago truncatula"
 /mol_type="mRNA"
 /cultivar="genotype A17"
 /db_xref="taxon:3880"
 /clone="PKV0-15G16"
 /tissue_type="Seedling roots"
 /dev_stage="Immediately prior to inoculation with
 Sinorhizobium meliloti (0 hour)"
 /lab_host="E.coli strain XL0LR"
 /clone_lib="KV0"
 /note="Vector: pBluescript SK -; Site 1: EcoRI; Site 2:
 XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
 was directionally ligated into the Unizap XR vector from

Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-Zap phage using Ex-assist
 helper phage and propagated in XL0LR cells."

ORIGIN

Query Match 9.5%; Score 398.8; DB 10; Length 610;
 Best Local Similarity 78.4%; Pred. No. 3.1e-79;
 Matches 478; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 1570 CGTGTGTGATCTGAGCAGTCTGCAGAAAGAAATATTACAAAGCTATTTTACCCGTATTAAT 1629
 Db 1 CGAGTGGATTTCAGCAGCAAAACAGAAAGATATTATAGCAATTTTACCCGTATTAAT 60
 QY 1630 CAAGTATTGACAAAAAAGGGAGGTGCTCAAAATTTCCCTTAAATAACATTAATGATGAATTA 1689
 Db 61 CAGATATTAAACCGCGCTGGTGTGCACAGATTTCTTTATCAACGTTGTTATGGAATG 120
 QY 1690 CGAAAAGTATGCTGCCATCTTTATATGCTAGAGGCTGTTGAGCCAGTTATTACGACGCA 1749
 Db 121 CGTAAAGCTCTGTGTGCTGCTTACATGTTAGAAAGAGTTGAGCCAGATATTACGATCCA 180
 QY 1750 AATGAAGCTTTCAAAACAACTTTTGGAGTCTTCTGGAAGAGCTGCAACTCTTAGATAAAATG 1809
 Db 181 AAAGAAGCATTCAGCAATTTGCTGGAATCATCAGGAAGTTGCATTTGCTTGACAAGATG 240
 QY 1810 ATGGTCAAACTGAAAGAGCAAGACACAGAGTCTTAATATACACAGATTTTCAGCATATG 1869
 Db 241 ATGGTGAAGCTTAAAGAACAGGACATAGAGTTCTCATATACTCCAGTTTCAGCACATG 300
 QY 1870 CTGGACTTACTTGAAGACTACTGTACCCATAGAATAAGCAGTACGAGCGAATTGATGGA 1929
 Db 301 CTTGATTTTCTTGAAGATTTACTGCTCTTCAAGAAATGCGCATTTATGAAGGATAGATGCG 360
 QY 1930 AAGGTTGGCGGAGCTGAGCGCAAAATACGATAGATCGTTTCAATGCCAAAAATTTCTAAC 1989
 Db 361 AAGTTTGGTGGGCTTGAAGACAAATACGATAGATCGTTTATGCGCAAAAATTTCTTCA 420
 QY 1990 AAGTTTGTGTTTTGCTCTTCCAAGAGCTGGTGGCTTAGGAATAAATCTTCAACGGCT 2049
 Db 421 AGATTTTGTCTTTCTTCTTCTACAGAGCGCGGGTGTGGGAATAAACCTTCAACTGCT 480
 QY 2050 GATACAGTAATCATTTATGACAGTCACTGGAATCTCATGCTGATCTTCAAGCAATGGCT 2109
 Db 481 GACACAGTTGTTATTTATGACAGTGAATGGAATCTCATGCTGACCTACAGCAATGGCT 540
 QY 2110 AGAGCTCATCGACTTGGCCAAACAAATAAGGTGATGATTTATAGGCTCATTAACCGAGC 2169
 Db 541 AGAGCTCATCGACTTGGCAAAACTAAACAGGTTGTTGATTTATAGGCTTATAACACGGGA 600
 QY 2170 ACCATTGAAG 2179
 Db 601 ACGATTGAAG 610

RESULT 7

CB635984

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

CB635984
 OSIIIEb16P12.f OSIIIEb Oryza sativa (indica cultivar-group) cDNA
 clone OSIIIEb16P12 5', mRNA sequence.

CB635984
 CB635984.1 GI:29630975
 EST.
 CB635984

Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 745)
 Jantauriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
 Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea

Unpublished (2003)
 Contact: Rod Wing
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu

PCR Primers
 FORWARD: gta aaa cga cgg cca gtg
 BACKWARD: gga aac agc tat gac cat g
 Plate: 16 row P column: 12
 Seq primer: gta aaa cga cgg cca gtg.

FEATURES
 source
 1. 745
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="IR36"
 /db_xref="taxon:39946"
 /clone="OSIIEb16P12"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSIIEb"
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (PO6-6-3)"
 XhoI; 24 hrs after inoculation with Rice Blast (PO6-6-3)"

ORIGIN
 Query Match 9.3%; Score 387.6; DB 14; Length 745;
 Best Local Similarity 72.9%; Pred. No. 1.1e-76;
 Matches 541; Conservative 0; Mismatches 194; Indels 7; Gaps 3;

Qy 1661 TTTCCCTTAATACATTATGATGAATACGAAAGTA-TGCTGCCATCTTATATGCTA 1719
 Db 4 TTTCACTATTTCATGAAGAATGGAACACTACTCAAACTATTGTTGCCATGCAATCATGACA 63

Qy 1720 GAGGGTGTGAGGCAGTTATTTCAGCAGCAAAATGAAGCTTTCAAACAACTTTTGGAGTCT 1779
 Db 64 GATGAACCGA--AGAGCCGCCAATTCAGAGAAGCTTTAGGAGGCTTTAGATCT 120

Qy 1780 TGTGAAAGCTGCAACTTCTAGATAAAATGATGTCMAAATGATGTCMAAAGCAAGCAGACAGA 1839
 Db 121 TCTGAAAATGGAGTCTTGACAAGATGATGTAAGCTGTAAGAGCAGGCTCAGG 180

Qy 1840 GTCCATAATACACAGTTTCAGCATATGCTGCACTTACTTGAAGTACTGTACCCAT 1899
 Db 181 GTTCTTATTTATTCACAGTTCCAGCACATGTTGACATCTTGAAGGATTATTTAAGCTAC 240

Qy 1900 AAGAAATGGCAGTACGAGCGAATTGATGGAAGGTTGGCGGAGCTGAGCGGCAATACGC 1959
 Db 241 CGAAATGGAGTTATGAGCGTATTGATGGAAATAGTGGTGTGAGGCGACATACGA 300

Qy 1960 ATAGATCGGTTCAATGCCAAAATCTTAACAAGTTTTTTTGTCTCTCCACAAGAGCT 2019
 Db 301 ATTGATCGTTCATGCTAAAATTTCTACTAGGTTTTTGTCTTCTTTCAACAGAGCT 360

Qy 2020 GGTGGCTTAGAATAAATCTGCAAGCTGATACAGTAATCATTTATGACAGTACGCTGG 2079
 Db 361 GGTGGCTTAGAATAAATCTGCAAGCTGATACAGTAATCATTTATGACAGTACGCTGG 420

Qy 2080 AATCCTCATGCTGATCTTCAAGCAATGGCTAGAGCTCATGACTTTGGCCAAACAAATAAG 2139
 Db 421 AATCCTCATGCTGATTTACAGCAATGGCTAGAGCTCATGACTTTGGCCAAACAAATAAG 480

Qy 2140 GTGATGATTTATAGGCTCATAAACCGAGGCCATTTGAAGAAAGGATGATGCAATTCAT 2199
 Db 481 GTTATGATATATAGGCTTTGTTAGTCGTTGGTACAAATTTGAGGAGCGGATGATGCACTTACA 540

Qy 2200 AAAAAGAAATGGTTCTAGACATCTTGTGTTGGGAAATCTC---AAAAACAAACATTT 2256
 Db 541 AAGAAAAAATGGTATTGGAGCACTTAGTTGTTGGCGCTCTCACGAAAGGCACTTAATATT 600

2257 AATCAGGAAGATTAGATGACATCATCAGGTATGATCAAGAGAGCTTTTTCGTAGTGA 2316
 Db 601 GTCCAGGAGAGTTGGATGATATTATTCGGCATGGCTCAAGAGAACTTTTTCGACGTGA 660

Qy 2317 GATGATGAGCAGGAAGCTCTGGAATAATTCATATGATGATGCGGCTATAGCAAAATG 2376
 Db 661 AATGATGAAGCCGGAATAATCTTGCCAAATCCATTATGATGATGCTGCGATTGATAAATTA 720

Qy 2377 CTTGATCGTGTATCTCGTGGAGG 2398
 Db 721 TTGGACCGTGACCAAGCTGAGC 742

RESULT 8
 AI998628/c 563 bp mRNA linear EST 08-SEP-1999
 LOCUS 701546371 A. thaliana, Columbia Col-0, rosette-2 Arabidopsis
 DEFINITION thaliana cDNA clone 701546371, mRNA sequence.
 ACCESSION AI998628
 VERSION AI998628.1 GI:5845533
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 563)
 AUTHORS Chen, J., Momiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D., Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P., Gorgone, G., Burns, D., Griffin, J., Mouanoutou, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S., Nobrega, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and Hanson, D.
 TITLE Arabidopsis thaliana Gene Expression MicroArray
 JOURNAL Unpublished (1999)
 COMMENT Contact: David Smoller, Ph.D.
 Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc.
 4633 World Parkway Circle, St. Louis, MO 63134, USA
 Tel: 877-577-2733
 Fax: 314-427-3324
 Email: service@genomesystems.com.
 Location/Qualifiers
 1. 563
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /cultivar="Columbia Col-0"
 /db_xref="taxon:3702"
 /clone="701546371"
 /tissue_type="rosette"
 /dev_stage="4 - 7 weeks"
 /clone_lib="A. thaliana, Columbia Col-0, rosette-2"
 /note="Vector: pSPORT; Site 1: NotI; Site 2: SalI; cDNA library was derived from untreated rosette tissue from Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. cDNA synthesis was initiated using a NotI-oligo(dT) primer. Double-stranded cDNA was bluntended, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and SalI sites of the pSPORT vector."

ORIGIN
 Query Match 9.2%; Score 385; DB 9; Length 563;
 Best Local Similarity 100.0%; Pred. No. 4.1e-76;
 Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3793 CAGATGTGCAAACTCTTGTATGAGAACGCTCGGGAATCAGTCCCAAGCATATGTAACAC 3852
 Db 563 CAGATGTGCAAACTCTTGTATGAGAACGCTCGGGAATCAGTCCCAAGCATATGTAACAC 504

3853 CAACCCGAGTACCAAGGTGAATGAGAGCTCCGTCGACTCAAACTATCAATGTTAAC 3912
 |||
 503 CAACCCGAGTACCAAGGTGAATGAGAGCTCCGTCGACTCAAACTATCAATGTTAAC 444
 |||
 3913 ATTAACACAATCTCTTTCGATTACATCTGATCAATCCAAAGTCACATGAAGACGACACCAAG 3972
 |||
 443 ATTAACACAATCTCTTTCGATTACATCTGATCAATCCAAAGTCACATGAAGACGACACCAAG 384
 |||
 3973 CAGACCTAAACAATGTTGAGATGAAGGACACCGCGGAGAAACAAACCCGTTAAGAGGT 4032
 |||
 383 CAGACCTAAACAATGTTGAGATGAAGGACACCGCGGAGAAACAAACCCGTTAAGAGGT 324
 |||
 4033 GCGGTGCTGCTGATCTGAATGTTGGTGGAGGAGGAGAACATCTCTGAAGCTAGTGAAGT 4092
 |||
 323 GCGGTGCTGCTGATCTGAATGTTGGTGGAGGAGGAGAACATCTCTGAAGCTAGTGAAGT 264
 |||
 4093 GTTGATGTTAAATGGAAGAACCAAGAGAGAGAGCAAGCAAGCAATGTCGTTGAT 4152
 |||
 263 GTTGATGTTAAATGGAAGAACCAAGAGAGAGAGCAAGCAAGCAATGTCGTTGAT 204
 |||
 4153 TGACTCAACTGGTAAATCAAGATTC 4177
 |||
 203 TGACTCAACTGGTAAATCAAGATTC 179
 |||

RESULT 9
 BE204298
 LOCUS
 DEFINITION BE204298 597 bp mRNA linear EST 05-SEP-2000
 EST396974 KVO Medicago truncatula cDNA clone pkV0-15G18, mRNA
 sequence.
 ACCESSION BE204298
 VERSION BE204298.1 GI:8747583
 KEYWORDS EST.
 SOURCE Medicago truncatula (barrel medic)
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.
 REFERENCE 1 (bases 1 to 597)
 VandenBosch, K., Endre, G., Hur, J., Moore, J., Beremand, P., Ellis, L.,
 Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and
 Fraser, C.M.
 ESTs from uninoculated seedling roots of Medicago truncatula
 Unpublished (1999)
 Contact: VandenBosch K
 Department of Plant Biology
 University of Minnesota
 220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
 Tel: 612 624 2755
 Fax: 612 625 1738
 Email: kvanden@cbs.umn.edu
 Texas A&M University name:T264070e
 TIGR sequence name:MTGAR45TK
 More information is available at:
 http://chryslr.tamu.edu/medicago
 Seq primer: Skmod (CTA gaa cta gtg gat CC).
 Location/Qualifiers
 1..597
 /organism="Medicago truncatula"
 /mol_type="mRNA"
 /cultivar="genotype A17"
 /db_xref="taxon:3880"
 /clone="pkV0-15G18"
 /tissue_type="Seedling roots"
 /dev_stage="Immediately prior to inoculation with
 Sinorhizobium meliloti (0 hour)"
 /lab_host="E.coli strain XLOLR"
 /clone_lib="KVO"
 /note="Vector: pBluescript SK -; Site 1: EcoRI; Site 2:
 XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
 was directionally ligated into the Unizap XR vector from
 Stragatene and packaged using Gigapack III Gold packaging

extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-Zap phage using Ex-assist
 helper phage and propagated in XLOLR cells."

ORIGIN
 Query Match 8.9%; Score 373; DB 10; Length 597;
 Best Local Similarity 77.9%; Pred. No. 2.1e-73;
 Matches 462; Conservative 0; Mismatches 130; Indels 1; Gaps 1;
 1570 CCGTGTGATCTCAGCAGCTCTGCAGAAAGAAATATTACAAAGCTATTTTACCCTGTAATAT 1629
 |||
 2 CCGAGTGGATTCTGAGCAGCAACAGAAAGAAATATTATAAGCAATTTTGACCCGTAATAT 61
 |||
 1630 CAAAGTATTACAAAAGAGGAGGTGCTCA-ANTTTCCCTTAATAACATTTATGATGAAT 1688
 |||
 62 CAGATATTAAACCCGCGGTGGTGCACAGTATTTCTTATCAACGTTGTTATGGAAT 121
 |||
 1689 ACGAAAGATGCTGCGCATCTTATATGCTAGAGGGTGTGAGCCAGTATTATTCACGACGC 1748
 |||
 122 GCGTAAAGCTCTGTTGTCATGCTTACATGTTAGAGGAGTTGAGCCNGATATTGACGATCC 181
 |||
 1749 AAATGAAGCTTTCAACCAACTTTTGGAGTCTTTGGAAAGCTGCAACTTCTAGATAAAT 1808
 |||
 182 AAAAGAAGCATTTCAAGCAATTCGTGATCATCAGGGAAGTTGCACTTCTGTCACAAGAT 241
 |||
 1809 GATGTCAAACTGAAGAGCAAGGACACAGAGCTCCTAATATACACACAGTTTTCAGCATAT 1868
 |||
 242 GATGTTGAAGCTTAAAGAACAGAGCATAGAGTTCTCATATATCTCCAGTTTTCAGCACAT 301
 |||
 1869 GCTGGACTTACTTGAAGACTACTGTACCCATGAAGAAATGGCAGTAGTCAGCGCAATTCATGG 1928
 |||
 302 GCTTGATTTGCTTGAAGATTTACTGCTCTTACAAGAAATGGCATTTATGAAGGATAGATGG 361
 |||
 1929 AAAGTTGGCGGAGCTGAGCGGCAAAATACGCATAGATCGTTCAATGCCAAAAATTCATA 1988
 |||
 362 CAAAGTTGCTGGGCTGAAAGACAAATACGGATAGATCGTTTAAATGCCAAAAATTCCTC 421
 |||
 1989 CAAAGTTTGTGTTTGTCTCTCCACAGAGCTGGTGGCTTAGGAATAAATCTTGCACACGC 2048
 |||
 422 AAGATTTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 481
 |||
 2049 TCATACAGTAATCAATTTATGACAGTGACTGGAAATCCTCATGCTGATCTTCTTCAAGCAATGGC 2108
 |||
 482 TCACACAGTTGTTATTTATGACAGGATTTGGAATCTCTCATGCTGACCTACAGCAATGGC 541
 |||
 2109 TAGAGCTCATCGACTTGGCCAAACAAATAAAGCTGATGATTTATAGCTCATAA 2161
 |||
 542 TAGAGCTCATCGACTTGGACAAACTAAACAAGGTGTTGATTTATAGCTTATAA 594
 |||

RESULT 10
 BQ139596
 LOCUS
 DEFINITION BQ139596 589 bp mRNA linear EST 26-APR-2002
 NF021H12PH1103 Phoma-infected Medicago truncatula cDNA clone
 NF021H12PH 5', mRNA sequence.
 ACCESSION BQ139596
 VERSION BQ139596.1 GI:20275722
 KEYWORDS EST.
 SOURCE Medicago truncatula (barrel medic)
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.
 REFERENCE 1 (bases 1 to 589)
 Watson, B.S., Shin, H.-S., Lopez-Meyer, M., Scott, A.D., Harris, A.R.,
 Gonzales, R.A., Bell, C.J., Inman, J.T., Waugh, M.E., Sullivan, J.P.,
 May, G.D. and Paiva, N.L.
 TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation
 JOURNAL Medicago truncatula Phoma-infected library
 COMMENT Unpublished (2002)
 Contact: Paiva NL
 Plant Biology Division

The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7317
Fax: 580 221 7380

Email: nlpaiva@noble.org
Insert Length: 589 Std Error: 0.00
Plate: 021 row: H column: 12
Seq primer: TCACACAGGAACAGCTATGAC.

FEATURES

Location/Qualifiers
1..589
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="NF021H12PH"
/tissue_type="leaf"
/dev_stage="Pathogen-induced, young trifoliolate"
/clone_lib="Phoma-infected"
/notes="Vector: pBluescript SK(-); Young trifoliolate leaves of Medicago truncatula were excised and dip-inoculated in a spore suspension of Phoma medicaginis, and incubated in humid dishes. Pools of leaves were harvested at 0, 15, and 30 minutes and 1, 2, 3, 6, 14, 24, 48, 72, and 96 hours, and used to prepare total RNA. cDNA was prepared from polyA+ enriched, pooled samples of equivalent amounts of total RNA from each sample. The cDNA was directionally ligated into the Uni-Zap XR vector (Stratagene) and packaged using the Gigapack III Gold packaging extracts. Phagemids containing cDNA inserts were in vivo excised from the recombinant Uni-Zap XR vector using ExAssist helper phage and the E. coli strain XL1-Blue MRF' (Stratagene). Excised plasmids were plated using SOLR cells."

ORIGIN

Query Match 8.5%; Score 354.6; DB 13; Length 589;
Best Local Similarity 77.6%; Pred. No. 3.1e-69;
Matches 456; Conservative 0; Mismatches 120; Indels 12; Gaps 2;

```
QY 919 CAAGAGTCCCTTTAGCTTACATTTTGGAGGAGCACTC-----ATTCCGCATTTC 972
DB 1 CAAAGTATTGCTTCTTAGCATCCCTTTTGAAGAGGCTGTCTCTGCACATCCACATCTG 60
QY 973 GTAATTCTCTCTATCGACTCGCTAACTGGGAGAGAGATTTGCCATATGGGCCCCA 1032
DB 61 GTGGTTGCTCCACTTCAACACTCGGAACCTGGGAAGTGAATTTGCAACATGGCCCT 120
QY 1033 CAGATGAACGTGGTATGATATTTGGCACTCGGCAAGCTCGAGAGTATTCAGAGAACAT 1092
DB 121 CAAATGAATGTTATTTATGTATGTTGATCTGCCCAAGCTCGTAGTGTATTCAGAGAATAT 180
QY 1093 GAGTTTACTATCGAAGATCAAAAAGATCAAGAAAGAAAGAAATCTGGCAAAATAGT 1152
DB 181 GAATTTTACTTCCCAAGAACTGAAGAAGCAAGAAAGAAAGAAATCT-----TTAGTT 234
QY 1153 AGCGAAAGCAAGCAAAAAGAAATCAAGTTTGATGTCCTCTCATCTGATGAGATGATC 1212
DB 235 AGTGAAGTAGCATGACAGATTAAGTTTGATGTCCTTTTGACATCATATGATGATC 294
QY 1213 AACTAGATTACAGAGTTCTAAACCAATTAAGTGGGAGTGCATGATGTTGATGAAGGT 1272
DB 295 AACTTAGACACACATCATTTAAACCTTATAAATGGGAGTGCATGATGATGATGAAGGT 354
QY 1273 CATCGACTGAATAATAGGATTCAAAGCTGCTCTCTTCATTGACACAGTATTCAGTTAAC 1332
DB 355 CACCGCTCAAAAATAGGATTCAAAATTAATTTCTTCATTGAAGCAATATCTTACCAGA 414
QY 1333 CACCGTATTTCTTGACAGGAAACACCACTTCAGAAACACTTTGGATGAACCTTTTCATGCTC 1392
DB 415 CATGCTGTCTTGTACTGGACTCTCTTCAGACAACTTTGGATGAACCTTTTATGCTT 474
QY 1393 ATGCAATTTCTTGATCGGGGAAGTTTGGAAAGTTTGGAGAGTTCCAGGAGGATTCAAA 1452
DB 475 ATGCAATTTCTTGATCGGGGAAGTTTGGAAAGTTTANAGGAATTTCCAAGAAGAATTCAAG 534
```

```
QY 1453 GATATTAAATCAAGAGGAGCAGATCTCAAGGTTGCACAAAATGTGGCTCC 1502
DB 535 GATATCAATCAAGAGGACAGATTTCAGAGCTTCATAAAATGCTGGCCCC 584
```

RESULT 11

BE205450
LOCUS BE205450
DEFINITION BE205450 583 bp mRNA linear EST 05-SEP-2000
ACCESSION BE205450
VERSION BE205450.1 GI:8748748
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula

REFERENCE

1 (bases 1 to 583)
VandenBosch,K., Endre,G., Hur,J., Moore,J., Beremand,P., Ellis,L., Town,C.D., Bowman,C.D., Craven,M.B., Hansen,T.S., Holt,I.B. and Fraser,C.M.

AUTHORS

ESTs from uninoculated seedling roots of Medicago truncatula

TITLE

JOURNAL

COMMENT

Contact: VandenBosch K
Department of Plant Biology
University of Minnesota
220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenb@cbs.umn.edu
Texas A&M University name:T265222e
TIGR sequence name:MTGBM73TK
More information is available at .
http://chryslie.tamu.edu/medicago
Seq primer: SKmod (CTA gAA CTA gtg gat CC).
Location/Qualifiers

FEATURES

source

1..583
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pKV0-21N1"
/tissue_type="Seedling roots"
/dev_stage="Immediately prior to inoculation with Sinorhizobium meliloti (0 hour)"
/lab_host="E.coli strain XL0LR"
/clone_lib="KV0"
/note="Vector: pBluescript SK -; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XL0LR cells."

ORIGIN

Query Match 8.4%; Score 351.4; DB 10; Length 583;
Best Local Similarity 77.4%; Pred. No. 1.6e-68;
Matches 456; Conservative 0; Mismatches 121; Indels 12; Gaps 2;

```
QY 914 CAATTCAAGCATGCCCCCTTTAGCTTCACTTTTGGAGGAGCACTC-----ATTCCGC 967
DB 1 CTATACAAAGTATGCTTCTTAGCATCCCTTTTGAAGAGGCTGTCTCTGCACATCCAC 60
QY 968 ATTTGTAATTGCTCTCTATCGACTCTCGTAACTGGGAGAGAGTTCACACATGGG 1027
DB 61 ATCTGTTGTTGCTCTCACTTTCAACACTGGAAACTGGGAACGTGAATTTGCAACATGGG 120
QY 1028 CCCACAGATGAACGTGGTATGATATTTGGCACTCGCAAGCTCGAGCAGTTATCAGAG 1087
```

121 CCCCTCAATGAATGTTTATTATGATGTTGATCTGCCAAGCTCGTAGTGTATCAGAG 180
Db
1088 AACATGAGTTTACTTATCGAAGATCAAAAAGATCAAGAAAAGAAATCTGGACAAA 1147
Qy
181 AATATGAATTTTACTTTCCCAAGAAATCGAAGAGAAACAAGAAAAGAAATCT-----T 234
Db
1148 TAAGTAGCGAAGCAAGCAAAAGAAATCAAGTTTGTATGCTCCCTCACATCGTATGAGA 1207
Qy
235 TAGTTAGTGAAGTGAACATGACAGGATTAAGTTTGTATGCTTTTGACATCATATGAGA 294
Db
1208 TGATCAACCTAGATTCACGAGTTCTTAAACCAATTAAGTGGAGTGCATGATTGTTGATG 1267
Qy
295 TGATCAACTTAGACACACATCATTTAAACCTATATAAATGGAGTGCATGATTGATG 354
Db
1268 AAGTGTACGACTGAAATAAAGATTCAAGCTGTTCTCTTCTCATGACACAGTATTCAA 1327
Qy
355 AAGGTCAACCCCTCAAAAATAAGGATTCAAAAATATTATTTCTTCAATGAAGCAATTTCTA 414
Db
1328 GTAAACACCGTATTCTTCTGACAGGAACACACATTCAGAACAACTTGGATGAACTTTTCA 1387
Qy
415 CCAGACATCGTGTGCTCTTGAATGGAATCTCTCTTCAAGAACACTTGGATGAACTTTTCA 474
Db
1388 TGCTCATGCAATTTCTTGTATGCGGGGAAGTTTGAAGTTTGGAGGAGTTCAGGAGGAGT 1447
Qy
475 TGCTTATGCAATTTCTTGTATGCGGGGAAGTTTGAAGTTTGGAGGAGTTCAGGAGGAGT 534
Db
1448 TCAAGATATTAATCAAGAGGAGCAGATCTCAAGTTTGGACAAATAATGTT 1496
Qy
535 TCAAGGATATCAATCAAGAGGAGCAGATTTCAAGCTTCATAAATGCT 583
Db

RESULT 12

CA240279 778 bp mRNA linear EST 25-SEP-2003
LOCUS SCSBFL4066A03.g FL4 Saccharum officinarum cDNA clone SCSBFL4066A03
DEFINITION 5', mRNA sequence.

ACCESSION CA240279.1 GI:35313723

VERSION CA240279.1

KEYWORDS EST.

SOURCE Saccharum officinarum

ORGANISM Saccharum officinarum

Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum.

1 (bases 1 to 778)

Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.

The libraries that made SUCST

Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

Contact: Arruda P

Centro de Biologia Molecular e Engenharia Genetica

Universidade Estadual de Campinas

Caixa Postal 6010, 13083-970, Campinas SP, Brazil

Tel: 55 19 3788 1137

Fax: 55 19 3788 1089

Email: parruda@unicamp.br

Clone distribution: clone distribution information can be found

through the Brazilian Clone Collection Center (BCCC) at

http://www.bcccenter.fcav.unesp.br

Plate: 066 row: A column: 03

Seq primer: T7 Promoter Primer.

Location/Qualifiers

1. 778

/organism="Saccharum officinarum"

/mol_type="mRNA"

/db_xref="taxon:4547"

/clone="SCSBFL4066A03"

/lab_host="DH10B"

/clone_lib="FL4"

/note="Organ: Developed inflorescence and rachis

(20cm-long); Vector: pSport1; Site 1: SalI; Site 2: NotI;

An unidirectional cDNA library generated from [Developed

inflorescence and rachis (20cm-long)]. cDNA was prepared

from polyA+ mRNA using Superscript Plasmid System Kit

(Invitrogen). The double-strand cDNAs were fractionated
in a sepharose CL-2B 40cm-columns and fragments sizing
between 0.8 and 1.5 Kb were directionally cloned into the
vector. Details of each source of RNA and library
construction can be obtained at
http://sucst.lad.ic.unicamp.br/public"

ORIGIN

Query Match 8.0%; Score 336; DB 14; Length 778;
Best Local Similarity 68.9%; Pred. No. 5.2e-65;
Matches 507; Conservative 0; Mismatches 214; Indels 15; Gaps 3;

Qy 1911 GTACGAGCGAATTTGATGGAAGGTTTCGCGAGCTGACGCGCAATACGCGATAGATCGGTT 1970
Db 55 GTCGGAACGATTTGATGGAAGGTTTCGCGAGCTGACGCGCAATACGCGATAGATCGGTT 114
Qy 1971 CAATGCCAAAAATTTCTAAACAAGTTTGTGTTTTCCTCTCCACAAAGAGCTGGTGGCTTAGG 2030
Db 115 CAATGCTAAGATTTCGACTAGGTTTGTCTTCTTCTACAGAGCTGGTGGCTGGG 174
Qy 2031 AATAAATCTTGCAACGGCTGATACGTAATCATTTATGACAGTGCATGGAATCCTCATGC 2090
Db 175 AATAAATTTGGCAACTGCAGATACTGTAATCATCTATGACAGTGTGGAACCCACATGC 234
Qy 2091 TGATCTTCAAGCAATCGCTAGAGCTCATCGACTTGGCCCAACAAATAAGTGATGATTTA 2150
Db 235 GGATTTGCAAGCTATGGCAAGAGCTCATCGCTTAGGACAGACTAGCAAGATGATGATATA 294
Qy 2151 TAGGCTCATAAACCCGAGGACCAATTGAAGAAGGATGATGCAATTCGCTTAAAAAGAAAT 2210
Db 295 CCGGCTTGTAGCCGAGGTACAATTGAGGNAAGATGATGCGCTTTACAAGAAAGAAAT 354
Qy 2211 GGTCTTAGAGCATCTTGTGTTGGGAACTC---AAAAACAACAAATTAATCAGGAAGA 2267
Db 355 TTATTGGAGCACTTAGTTGTTGGTGCAGCTCAGAAAAGCTAAATAATGTCATCAGGAGGA 414
Qy 2268 GTTAGATGACATCATCAGGTATGGATCAAGAGGCTTTTGTCTAGTGAAGATGATGAGC 2327
Db 415 GCTGGATGATATATACGCTTGGATCAAGAGGAGCTTTTGTATGACGAGAAATGACGAA-- 472
Qy 2328 AGAAAGTCTGAAAAAATTCATTATGATGATGCGGCTATAGACAAATTTGCTGATCGTGA 2387
Db 473 -----TCTGCCAAATTCATTACGCAAGCTGCAATTTGAGAGGTTGTTAGACGCTGA 535
Qy 2388 TCTCGTGGAGGAGGAGGAAGTCTCAGTGAATGAAGAGGAGAAATGGAATCTTTAAAGGC 2447
Db 526 TCAAGTTGATGGTGATGAA---TCTGTGAAGATGAAGAAGAAGATGAAATCTTTAAAGG 582
Qy 2448 TTTCAAGTGGCTAATTTTGAATATATAGATGAAATGAGGAGGAGCAGCATTAGAGGCACA 2507
Db 583 ATTCAAGGTTGCAAACTTTGAATATACATCGACGAGGAGGAGCTCANGCAGATAGAGGGA 642
Qy 2508 GAGAGTCGCTGTGAAAGCAAAATCTTCAGCAGGCAATTTCTGATAGAGCAAGTTATTGGGA 2567
Db 643 GGAGGCCCGCAGAAAGGCTGCAGCTGAGCTGAAAATCTCTGAAGATTAACCTATTGGGA 702
Qy 2568 AGAGTTGTTTAAAGATAAATTTGAGCTGCCAGGCTGAGGAGCTTAATGCTCTTGGAAA 2627
Db 703 TGAACATATGAAGGATAGATACGATGTACAGAAAAGTTGAGGAACATCTGCTATGGGAAA 762
Qy 2628 AAGGAGAGAAAGTCCG 2643
Db 763 AAGGAGAGAAAGACCCGC 778

RESULT 13

BQ834228

LOCUS

DEFINITION

Arabisidopsis lyrata cDNA clone p1WB2-E11 3', mRNA sequence.

ACCESSION

BQ834228.1

VERSION

KEYWORDS

BQ834228 348 bp mRNA linear EST 14-MAR-2003
Arabisidopsis lyrata cDNA clone p1WB2-E11 3', mRNA sequence.
Arabisidopsis lyrata cDNA clone p1WB2-E11 3', mRNA sequence.
BQ834228.1 GI:28951543
EST.

```
SOURCE Arabidopsis lyrata
ORGANISM Arabidopsis lyrata
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
TITLE rosid; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
JOURNAL 1 (bases 1 to 348)
MEDLINE Barrier,M., Bustamante,C.D., Yu,J. and Purugganan,M.D.
PUBMED Selection on rapidly evolving proteins in the Arabidopsis genome
22505405 Genetics 163 (2), 723-733 (2003)
COMMENT 12618409
Contact: Barrier M
Department of Genetics
North Carolina State University
3513 Gardner Hall, Box 7614, Raleigh, NC 27695, USA
Tel: 919 515 1761
Fax: 919 515 1695
Email: mbarrie@unity.ncsu.edu
Plate: 1 row: J column: 22
Seq primer: T3.
FEATURES             Location/Qualifiers
     source            1..348
     organism="Arabidopsis lyrata"
     mol_type="mRNA"
     cultivar="Karhumaki"
     db_xref="taxon:59689"
     clone="PIWB2-E11"
     tissue_type="Inflorescence"
     clone_lib="Arabidopsis lyrata Inflorescence pCMV-PCR
     Library"
     note="Vector: pCMV-PCR (Stratagene); Created using PCR
     Library Construction kit (Stratagene)"
ORIGIN
Query Match       7.4%; Score 308; DB 13; Length 348;
Best Local Similarity 92.8%; Pred. No..9.8e-59;
Matches 323; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 3657 GGGAGTGAGCTTATTGAGGTGTGATGTAATGCTGTGAGCTTCTTAAGACTGATCC 3716
Db 1 GGGAGTGAGCTTATTGAGGTGTGATGTAATGCTGTGAGCTTCTTAAGACTGATCC 60
QY 3717 TATCACTTTCAGAGAAATATTGGGGGCTGCTGTGTGACAAACCAACGCGGGTGGAAAT 3776
Db 61 CATCACTTCAGAGAAATATTGGTGGCTGTGTGACAAACCAACGCGGGTGGAAAT 120
QY 3777 AGCTCAACATTATTAACAGATGTGCAAACTTCTTGATGAGAACGCTCGGGAATCAGTCCA 3836
Db 121 AGCTCAACATTATTAACAGATGTGCAAAAGTTTTTAATGAGAACGCTCGTGAATCACTCCA 180
QY 3837 AGCATATCTTAACACCAACCAACCAAGTACCAAGTGAATGAGAGCTTCCGTGCACTCAA 3896
Db 181 AGCATATCTTAACACCAACCAACCAAGTACCAAGTGAATGAGAGCTTCTGTGCACTCGA 240
QY 3897 ATCTATCAATGTTAACTTAACAAATCCTTTTCGATTACATCTGATCAATCAAGTCCACA 3956
Db 241 ATCTATCAATGTTAACTTAACAAATCCTTTTCGATTACATCTGATCAATCAAGTCCACA 300
QY 3957 TGAAGACGACACCAAGCCAGACCTTAACAAATGTTGAGATGAAGGACAC 4004
Db 301 TGAAAACGACACCAAGCCAAACCTTAACAAATGTTGACATGAAGGACAC 348
RESULT 14
B29916/c 503 bp DNA linear GSS 13-OCT-1997
LOCUS T19D20RFB TAMU Arabidopsis thaliana genomic clone T19D20, genomic
DEFINITION survey sequence.
ACCESSION B29916
VERSION B29916.1 GI:2515882
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 503)
Rounsley,S.D., Kelley,J.M., Field,C.E., Craven,M.B., Adams,M.D. and
Venter,J.C.
Use of a BAC End Sequence Database To Identify Minimal Overlaps for
Arabidopsis Genomic Sequencing
Unpublished (1997)
Other GSSs: T19D20TFB
Contact: Steve Rounsley
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: rounsley@tigr.org
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 503.
FEATURES             Location/Qualifiers
     source            1..503
     organism="Arabidopsis thaliana"
     mol_type="genomic DNA"
     strain="Columbia"
     db_xref="taxon:3702"
     clone="T19D20"
     sex="hermaphrodite"
     clone_lib="TAMU"
     notes="Vector: BelOBACII; Site_1: HindIII; Site_2:
     HindIII; Produced by Rod Wing"
ORIGIN
Query Match       7.2%; Score 300; DB 28; Length 503;
Best Local Similarity 100.0%; Pred. No.6.7e-57;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3878 AGAGCTTCCTGTCACCTCAATCTATCAATGTTAATCAATTAACCAATCTTCGATTACAT 3937
Db 503 AGAGCTTCCTGTCACCTCAATCTATCAATGTTAATCAATTAACCAATCTTCGATTACAT 444
QY 3938 CTGATCAATCAAGTCATCAATGAAGACGACACCAAGCCAGACCAATGTTGAGATGA 3997
Db 443 CTGATCAATCAAGTCATCAATGAAGACGACACCAAGCCAGACCAATGTTGAGATGA 384
QY 3998 AGGACACGCGCCGAGAAACAAACCGTTAAGAGTGGCGTCGATCTCAATGTTGGTGG 4057
Db 383 AGGACACGCGCCGAGAAACAAACCGTTAAGAGTGGCGTCGATCTCAATGTTGGTGG 324
QY 4058 AGGAGAGGAGAACATTCGTGAAGCTAGTGAAGTGTGATGTAAATAATGGAAGAGCCA 4117
Db 323 AGGAGAGGAGAACATTCGTGAAGCTAGTGAAGTGTGATGTAAATAATGGAAGAGCCA 264
QY 4118 AGAAGAAGAGAGCCAAAGAACATGTGCTGTGATGACCTCAACTGTAATCAAGATTTC 4177
Db 263 AAGAAGAAGAGAGCCAAAGAACATGTGCTGTGATGACCTCAACTGTAATCAAGATTTC 204
RESULT 15
CD812910 673 bp mRNA linear EST 10-JUL-2003
LOCUS BN10.022P15F020211 BN10 Brassica napus cDNA clone BN10022P15, mRNA
DEFINITION sequence.
ACCESSION CD812910
VERSION CD812910.1 GI:32494850
KEYWORDS Brassica napus (rape)
SOURCE Brassica napus
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 673)
REFERENCE
AUTHORS Genopiante.
```

TITLE Genoplate, a major partnership french program in plant genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Genoplate
Genoplate
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10

This sequence has been generated in the framework of the french
plant genomics programme 'Genoplate' (<http://www.genoplate.com>
and <http://genoplate-info.infobiogen.fr>).

FEATURES

source
1..673
Location/Qualifiers
/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="Jet neuf"
/db_xref="taxon:3708"
/clone="BN1002P15"
/tissue type="seed"
/clone_lib="BN10"

ORIGIN

Query Match 7.1%; Score 298.2; DB 14; Length 673;
Best Local Similarity 81.5%; Pred.No.1.8e-56;
Matches 400; Conservative 0; Mismatches 73; Indels 18; Gaps 4;
QY 3642 GATCGCTGACACAGTGGGAGTGAGCTTTATTGAGGTTGATGATGAATGCTTGATGACT 3701
Db 1 GGTCACTGACACAGTGGGGCCAGCTTTATTGAGGTCGATGATGAATGCTTCAGGCAC 60
QY 3702 TCCTAAGACTGATCCTATCACTTCAGAAGAAATTTATGGGGGCTGCTTGACAAACCA 3761
Db 61 TCCCAAGACTGAGCCCATCACTTCAGAAGAGATTATGGTTGCTGCTGACAAACCA 120
QY 3762 AGCGCGGTCGAATAGCTCAACATTATTAACAGATGTGCAAACTTCTTGATGAGAACGC 3821
Db 121 AGAGCGGTTGAGTAGCCCACTTTATAACAAGATGTGCAAGATTTAGATGAGAACGC 180
QY 3822 TCGGGAATCAGTCCAGCATATGTAAACCAACCAACCGAGTACCAGGTGAATGAGAG 3881
Db 181 TCGTGAATCTGTCCAAGCATATGTGAACCAACCAACCGTCTAATGGCAAACTGGGCGAGAG 240
QY 3882 CTTCCGTCGACTCAAACTATCAATGGTAACATTAAACAATCCTTTTCGATTACATCTGA 3941
Db 241 CTTCCGTTCCATAGATC-----TACATTAGAGAAATCCTTTCCGCTCCATCTGA 291
QY 3942 TCAATCCAAGTCATGAAGACGACACCAAGCCAGACCTAAACAATGTTGAGATGAAGGA 4001
Db 292 TCAACTCAAGTCTACTGAA---GACACCAAGCCGACCTAAACAATGTTGAGAT---GGA 345
QY 4002 CACGCGCGAAGAAACAAACCGTTAAGAGTGGCGTCTGCTGATCTGAATGTTGAGAGGG 4061
Db 346 CTCGGTTCGAAGAAACGAAACCCCTTGAGAGCAAG---CGTCGATCTGAATGTTGAGAGGG 402
QY 4062 AGAGGAGAACATGCTGAAGCTAGTGGAGTGTGATGATTAATAATGGAGAGCCCAAGA 4121
Db 403 AGAGGAGAACATTTGGAGAGGTTAGTGAAGTGTGATGATGAAGATGAAGAGCCCAATG 462
QY 4122 AGAAGAGAGC 4132
Db 463 AGAAGAGAGC 473

Search completed: September 18, 2004, 23:18:38
Job time : 9901 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 14, 2004, 01:51:05 ; Search time 158 Seconds
(without alignments)
2474.973 Million cell updates/sec

Title: US-10-049-137-2
Perfect score: 7187
Sequence: 1 MSSLVRLRIRSDKPYNL.....VDVKMEAEKPKMNVVD 1384

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7187	100.0	1384	4 AAY72755	Aay72755 Arabidops
2	7184	100.0	1384	4 AAY72756	Aay72756 Arabidops
3	7011.5	97.6	1357	4 AAY72787	Aay72787 Arabidops
4	2091	29.1	1912	6 AAE36110	Aae36110 Human chr
5	2091	29.1	1912	6 AAE36109	Aae36109 Human chr
6	2088	29.1	1911	2 AAR99534	Aar99534 Dermatomy
7	2083	29.0	1912	4 AAM39288	Aam39288 Human pol
8	2083	29.0	1944	6 AAE36107	Aae36107 Human chr
9	2067.5	28.8	1936	4 AAM41074	Aam41074 Human pol
10	2007	27.9	1470	5 ABP64883	Abp64883 Human pro
11	1939.5	27.0	1377	6 AAE36108	Aae36108 Human chr
12	1874.5	26.1	979	7 ADB64292	Adb64292 Human pro
13	1738.5	24.2	883	4 ABB71771	Abb71771 Drosophil
14	1621.5	22.6	1739	6 AAE36105	Aae36105 Human chr
15	1621.5	22.6	1739	6 AAE36106	Aae36106 Human chr
16	1665.5	21.8	1559	6 AAE36103	Aae36103 Human chr
17	1562.5	21.7	1709	6 AAE36104	Aae36104 Human chr
18	1562.5	21.7	1709	6 ABO07199	Abo07199 Human p53
19	1522.5	21.2	1883	4 ABB60038	Abb60038 Drosophil
20	1498	20.8	2703	4 ABB60074	Abb60074 Drosophil
21	1492.5	20.8	1967	6 AAE35342	Aae35342 Human col
22	1492.5	20.8	1967	6 AAE36114	Aae36114 Human chr
23	1461.5	20.3	1468	6 ABR52904	Abr52904 Protein s
24	1422.5	20.1	2897	6 ABR82557	Abr82557 Human chr
25	1422.5	19.8	2713	6 AAE36113	Aae36113 Human chr

26	1407.5	19.6	2131	6 AAE36112	Aae36112 Human chr
27	1374.5	19.1	1010	4 AEG09719	Abg09719 Novel hum
28	1260	17.5	2464	4 AAM78919	Aam78919 Human pro
29	1224.5	17.0	1057	5 AAU93152	Aau93152 Arabidops
30	1224.5	17.0	1057	7 ADD31011	Add31011 Plant yie
31	1157	16.1	730	4 AAB94089	Aab94089 Human pro
32	1156	16.1	730	5 ABB85037	Abb85037 Pain regu
33	1146.5	16.0	655	4 ABO09856	Abg09856 Novel hum
34	1139	15.8	1120	6 ABR52907	Abr52907 Protein s
35	1135.5	15.8	1129	6 ABR52906	Abr52906 Protein s
36	1131	15.7	1027	4 ABB63954	Abb63954 Drosophil
37	1131	15.7	1027	7 ADB79883	Adb79883 Rat putat
38	1120	15.6	1359	6 ABR53192	Abr53192 Protein s
39	1097	15.3	1302	5 ABP73697	Abp73697 Candida a
40	1095.5	15.2	1703	6 ABR52985	Abr52985 Protein s
41	1086.5	15.1	1104	6 ABJ25604	Abj25604 Aspergill
42	1086.5	15.1	1353	6 ABJ26204	Abj26204 Aspergill
43	1040.5	14.5	1537	4 ABB66989	Abb66989 Drosophil
44	1040.5	14.5	1638	4 ABB61946	Abb61946 Drosophil
45	1039.5	14.5	551	5 ABB98113	Abb98113 CHD activ

ALIGNMENTS

RESULT 1
AAY72755

ID AAY72755 standard; protein; 1384 AA.

XX AAY72755;

XX 31-MAY-2001 (first entry)

DE Arabidopsis thaliana PKL protein.

XX PKL: chromo domain; helicase domain; DNA binding domain; CHD3 protein;
XX zinc finger domain; transgenic plant; chromatin remodelling factor; LEC1;
XX pickle root; developmental identity; embryo development.

OS Arabidopsis thaliana.

Key Location/Qualifiers

Domain /label= zinc_finger_domain

FT /note= "This region is specifically referred in in claim

FT 8"

FT Domain 115..151

FT /label= Chromo domain_I

FT /note= "This region is specifically referred in in claim

FT 7"

FT Domain 191..227

FT /label= Chromo domain_II

FT /note= "This region is specifically claimed in in claim

FT 9"

FT Domain 293..739

FT /label= Helicase_domain

FT /note= "This region is specifically claimed in in claim

FT 7"

FT Binding-site 1069..1095

FT /label= DNA binding_domain

FT /note= "This region is specifically claimed in in claim

FT 7"

WO200114519-A2.

PD 01-MAR-2001.

XX 18-AUG-2000; 2000WO-US022725.

XX 20-AUG-1999; 99US-0149975P.

PA (PURD) PURDUE RES FOUND.

XX

PI	Ogas JP, Somerville CR;	Qy	601	DKWVKLKEQHRVLIYTOFQHMLDLLEDYCTHKQYERIDGKVGGAERQIRIDR	NAK	660
XX	WPI; 2001-218433/22.	Db	601	DKWVKLKEQHRVLIYTOFQHMLDLLEDYCTHKQYERIDGKVGGAERQIRIDR	NAK	660
DR	N-PSDB; AAD02817.	Qy	661	NSKFCFLSTRAGGGINLATATDTWIIYDSDNWPHADLOAMARHLGOTNKWIVRLI		720
XX	Regulating plant developmental identity comprises transforming a host	Db	661	NSKFCFLSTRAGGGINLATATDTWIIYDSDNWPHADLOAMARHLGOTNKWIVRLI		720
PT	cell with a nucleic acid encoding a PKL protein having at least one	Qy	721	NRGTIERMMQLTKKQWVLEHLVVGKLTQINIQEELDDIIRYGSKELFASEDEACKSG		780
PT	chromo domain, a helicase domain and a DNA binding domain.	Db	721	NRGTIERMMQLTKKQWVLEHLVVGKLTQINIQEELDDIIRYGSKELFASEDEACKSG		780
XX	Claim 62; Page 67-76; 87pp; English.	Qy	781	KIHYDDAIDKLDRDLVEAEVSVDEEENGFLKAFKVFANFEYIDNEAAALEAQRVAA		840
CC	The invention relates to a method of regulating plant developmental	Db	781	KIHYDDAIDKLDRDLVEAEVSVDEEENGFLKAFKVFANFEYIDNEAAALEAQRVAA		840
CC	identity which involves transforming a host cell with a nucleic acid	Qy	841	ESKSSAGNSDRASYWEELLKDKFELHQAELNALGKRKRKQIOLVSEEDDLAGLEDVSS		900
CC	encoding PKL (pickle) protein having at least one chromo domain, a	Db	841	ESKSSAGNSDRASYWEELLKDKFELHQAELNALGKRKRKQIOLVSEEDDLAGLEDVSS		900
CC	helicase domain, a DNA binding domain and a zinc finger domain. This	Qy	901	DGDESYEAESTDCEAAGQVOTGRRPYRRKGRONLEPTPLMEGEGRSFVLFNFQSORAI		960
CC	method is useful for transforming a host cell with PKL proteins for	Db	901	DGDESYEAESTDCEAAGQVOTGRRPYRRKGRONLEPTPLMEGEGRSFVLFNFQSORAI		960
CC	regulating developmental identity, such as for regulating the transition	Qy	961	FVQTLNRYGAGNFDWKEFVPRLKQKTFEINEXGILFLKHAIEIDENSPTFSDGVPKEG		1020
CC	from embryonic to post-embryonic development. The PKL proteins act as	Db	961	FVQTLNRYGAGNFDWKEFVPRLKQKTFEINEXGILFLKHAIEIDENSPTFSDGVPKEG		1020
CC	chromatin remodelling factors to repress transcription of LEC1, a protein	Qy	1021	LRIEDVLVRIALLILVQEKVVFEDHGPVFPFSRILRERPPGLRSGKIWEEDHKIMIRA		1080
CC	that plays a role in regulating embryo development. The method is also	Db	1021	LRIEDVLVRIALLILVQEKVVFEDHGPVFPFSRILRERPPGLRSGKIWEEDHKIMIRA		1080
CC	used for generating transgenic plants expressing PKL proteins. The	Qy	1081	VLKHGGRWQAI VDDKELGIGQELICKELNPPHISLSAAEQAGLQGGSGSNPGAQTNQ		1140
CC	present sequence is Arabidopsis thaliana PKL protein found in pickle	Db	1081	VLKHGGRWQAI VDDKELGIGQELICKELNPPHISLSAAEQAGLQGGSGSNPGAQTNQ		1140
CC	root. PKL functions in regulating developmental identity in host cells.	Qy	1141	NPGSVITGNNSADGAQVNSMFYRDMQRRLVEFVKRVLLEKAMNYEYAEYYGLGG		1200
CC	It is classified as CHD3 protein based on the presence of chromo domain,	Db	1141	NPGSVITGNNSADGAQVNSMFYRDMQRRLVEFVKRVLLEKAMNYEYAEYYGLGG		1200
CC	SNF2-related helicase/ATPase domain, DNA-binding domain and PHD zinc	Qy	1201	SSSIPTPEEPAEKIADTVGVSVIEVDDEMDLGLPKTDPITTSBEIMGAAVDNNQARVEIA		1260
CC	finger domain	Db	1201	SSSIPTPEEPAEKIADTVGVSVIEVDDEMDLGLPKTDPITTSBEIMGAAVDNNQARVEIA		1260
XX	Sequence 1384 AA;	Qy	1261	QHYNQMKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNINTILSITSDQSKSHE		1320
Qy	Query Match 100.0%; Score 7187; DB 4; Length 1384;	Db	1261	QHYNQMKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNINTILSITSDQSKSHE		1320
Best Local Similarity 100.0%; Pred. No. 0;		Qy	1321	DDTKPOLNNVEMKDTAETKPLRGVVVDLNVVGEENIAEASGVVDVVMEEAEKEEPKN		1380
Matches 1384; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Db	1321	DDTKPOLNNVEMKDTAETKPLRGVVVDLNVVGEENIAEASGVVDVVMEEAEKEEPKN		1380
1 MSSLVERLRISDRKPVYVNLDDSDDDFVPPKDTFEQVEAIVETDAKACQAGSESTN 60		Qy	1381	MVVD 1384		
1 MSSLVERLRISDRKPVYVNLDDSDDDFVPPKDTFEQVEAIVETDAKACQAGSESTN 60		Db	1381	MVVD 1384		
61 LVSCNCTCYAFAKLVPPPLKDAVENWRCPECVSPLENDKILDCEMRPTKSSEQSSD 120		Qy	RESULT 2			
61 LVSCNCTCYAFAKLVPPPLKDAVENWRCPECVSPLENDKILDCEMRPTKSSEQSSD 120		Db	AA72756			
121 AEPKPIFVKQVLVWKVGLSYLHCSWPEKEPQKAYKSNHRLKTRVNNFHRQMESFNNSD 180		Qy	ID AAY72756 standard; protein; 1384 AA.			
121 AEPKPIFVKQVLVWKVGLSYLHCSWPEKEPQKAYKSNHRLKTRVNNFHRQMESFNNSD 180		Db	XX	AA72756;		
181 DFVAIRPEWTVDRILACREEDGELEVLYVKYKELSYDECYWESESDISTPONEIQRFKDV 240		Qy	XX	AC		
181 DFVAIRPEWTVDRILACREEDGELEVLYVKYKELSYDECYWESESDISTPONEIQRFKDV 240		Db	DT	31-MAY-2001 (first entry)		
241 NSRTRRSKDVDRKRNPRDFQFDHTPEFLKGLLHPYQLEGFLNFRFSWSKQTHVILADEM 300		Qy	DE	Arabidopsis thaliana mutant PKL protein (K304R).		
241 NSRTRRSKDVDRKRNPRDFQFDHTPEFLKGLLHPYQLEGFLNFRFSWSKQTHVILADEM 300		Db	KW	PKL; chromo domain; helicase domain; DNA binding domain;		
301 GLGKTIQSIALLASLFEENLIPHLVIAPLSTLRNWEREFATWAPQMNVMVYFGTAQARAV 360		Qy	KW	zinc finger domain; transgenic plant; chromatin remodelling factor; LEC1;		
301 GLGKTIQSIALLASLFEENLIPHLVIAPLSTLRNWEREFATWAPQMNVMVYFGTAQARAV 360		Db	KW	mutant; mutein; pickle; developmental identity; embryo development.		
361 IREHEFYLSKDOKIKKKKSGQISSESQKQIKFEDVLLTSYEMINLDSAVLKPWKQCM 420		Qy	OS	Arabidopsis thaliana.		
361 IREHEFYLSKDOKIKKKKSGQISSESQKQIKFEDVLLTSYEMINLDSAVLKPWKQCM 420		Db	OS	Synthetic.		
421 VDEGHLRKNKDSKLFSSLLTOYSSNHRILLTGTPQNQLDELFLMLMHLFDAGKFGSLEBFQ 480		Qy				
421 VDEGHLRKNKDSKLFSSLLTOYSSNHRILLTGTPQNQLDELFLMLMHLFDAGKFGSLEBFQ 480		Db				
481 BEFKDINOEBQISRLHKMLAPHLRLRRVKQVWKMPPPKKELILRLVDLSLQKYYKAIPT 540		Qy				
481 BEFKDINOEBQISRLHKMLAPHLRLRRVKQVWKMPPPKKELILRLVDLSLQKYYKAIPT 540		Db				
541 RNYQVLTCKGAQOISLNNIMWELRKVCCHPYMLGVEPVIHDANEAFKQLLESCKQLQLL 600		Qy				
541 RNYQVLTCKGAQOISLNNIMWELRKVCCHPYMLGVEPVIHDANEAFKQLLESCKQLQLL 600		Db				

XX	Key	Location/Qualifiers	XX
XX	FT	Misc-difference 304	XX
XX	FT	/note= "Wild type Lys substituted by Arg"	XX
XX	FT		XX
XX	PN	W0200114519-A2.	XX
XX	PD	01-MAR-2001.	XX
XX	XX		XX
XX	PF	18-AUG-2000; 2000WO-US022725.	XX
XX	PR	20-AUG-1999; 99US-0149975P.	XX
XX	PA	(PURD) PURDUE RES FOUND.	XX
XX	XX		XX
XX	PI	Ogas JP, Somerville CR;	XX
XX	PI	WPI; 2001-218433/22.	XX
XX	DR	N-PSDB; AAD02817.	XX
XX	PT	Regulating plant developmental identity comprises transforming a host	XX
XX	PT	cell with a nucleic acid encoding a PKL protein having at least one	XX
XX	PT	chromo domain, a helicase domain and a DNA binding domain.	XX
XX	XX		XX
XX	PS	Example 4; Page; 87pp; English.	XX
XX	CC	The invention relates to a method of regulating plant developmental	XX
XX	CC	identity which involves transforming a host cell with a nucleic acid	XX
XX	CC	encoding PKL (pickle) protein having at least one chromo domain, a	XX
XX	CC	helicase domain, a DNA binding domain and a zinc finger domain. This	XX
XX	CC	method is useful for transforming a host cell with PKL proteins for	XX
XX	CC	regulating developmental identity, such as for regulating the transition	XX
XX	CC	from embryonic to post-embryonic development. The PKL proteins act as	XX
XX	CC	chromatin remodelling factors to repress transcription of LEC1, a protein	XX
XX	CC	that plays a role in regulating embryo development. The method is also	XX
XX	CC	used for generating transgenic plants expressing PKL proteins. The	XX
XX	CC	present sequence is Arabidopsis thaliana mutant PKL protein (K304R). PKL	XX
XX	CC	functions in regulating developmental identity in host cells. Note: the	XX
XX	CC	present sequence is not shown in the specification, but is derived from	XX
XX	CC	the Arabidopsis thaliana PKL sequence given in the sequence listing	XX
XX	CC	(RAY7255)	XX
XX	XX		XX
XX	SQ	Sequence 1384 AA;	XX
XX	Query Match	100.0%; Score 7184; DB 4; Length 1384;	XX
XX	Best Local Similarity	99.9%; Pred. No. 0;	XX
XX	Matches 1383; Conservative	1; Mismatches 0; Indels 0; Gaps 0;	XX
Qy	1	MSSILVERLRIRSRKPVNLDSDDDDDFFPKKDRTFEQVEAIVRTADKENACQACGSGSTN 60	Qy
Db	1	MSSILVERLRIRSRKPVNLDSDDDDDFFPKKDRTFEQVEAIVRTADKENACQACGSGSTN 60	Db
Qy	61	LVSNCNTCTAFAPKALVPPLKQASVENWRCPCVSPLEIDKILDCENRPTKSGQSSD 120	Qy
Db	61	LVSNCNTCTAFAPKALVPPLKQASVENWRCPCVSPLEIDKILDCENRPTKSGQSSD 120	Db
Qy	121	AEPKPIFVKQYLVKMGSLYLHCSWVPEKEFOKAYKSNHRLKTRVNNFHRQWESFNNSD 180	Qy
Db	121	AEPKPIFVKQYLVKMGSLYLHCSWVPEKEFOKAYKSNHRLKTRVNNFHRQWESFNNSD 180	Db
Qy	181	DFVAIRPEWTTVDRIILACREEDGELEYLVKYKELSYDECYWESESDISTFQNEIQRFKDV 240	Qy
Db	181	DFVAIRPEWTTVDRIILACREEDGELEYLVKYKELSYDECYWESESDISTFQNEIQRFKDV 240	Db
Qy	241	NSRTRRSKVDHKNRPDRFQPDHTPEFLKGLLHPYQLEGLNFIKLSKQTHVILADEM 300	Qy
Db	241	NSRTRRSKVDHKNRPDRFQPDHTPEFLKGLLHPYQLEGLNFIKLSKQTHVILADEM 300	Db
Qy	301	GLGRTIQSIALLASLFEENLIPLHVIAPLSTLRNWEREFATWAPQNMVVMYFGTAQARAV 360	Qy
Db	301	GLGRTIQSIALLASLFEENLIPLHVIAPLSTLRNWEREFATWAPQNMVVMYFGTAQARAV 360	Db
Qy	361	IREHEFYLSKQDKKIKKKSGQISSESKQRIKFDVLLTSYEMINLDSAVLKPKECMI 420	Qy

Db 1174 SSSIPTPEAPKADTVGVSVFVDEMDLGLPKTDPITSEIMGAANNQARVEIA 1233
Qy 1261 QHYNOMCKLLDENARESVOYVNNQPPSTKVNBSFRALKSINGNINTILSDQSKSHE 1320
Db 1234 QHYNOMCKLLDENARESVOYVNNQPPSTKVNBSFRALKSINGNINTILSDQSKSHE 1293
Qy 1321 DDTKPDLLNNVEMKDTAETKPLRGVVDLNNVSGEENIAASGVDDVMEEAEKPKN 1380
Db 1294 DDTKPDLLNNVEMKDTAETKPLRGVVDLNNVSGEENIAASGVDDVMEEAEKPKN 1353
Qy 1381 MVVD 1384
Db 1354 MVVD 1357

RESULT 4
AAE36110
ID AAE36110 standard; protein; 1912 AA.
XX AAE36110;
XX 26-JUN-2003 (first entry)
XX Human chromodomain helicase DNA binding protein (CHD) #8.
XX Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;
KW chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;
KW cell proliferative disorder; chromatin organisation modifier domain;
KW cytosolic.
XX Homo sapiens.
XX Key Location/Qualifiers
FH Domain 619..661
FT Domain /note= "Chromodomain"
FT Domain 729..1025
FT Domain /note= "SNF2N domain"
FT Domain 1090..1164
FT Domain /note= "Conserved C-terminal domain"
XX WO200298899-A2.
XX 12-DEC-2002.
XX 03-JUN-2002; 2002WO-US017466.
XX 05-JUN-2001; 2001US-0296076P.
PR 10-OCT-2001; 2001US-0328605P.
PR 22-OCT-2001; 2001US-0338733P.
PR 15-FEB-2002; 2002US-0357253P.
PR 15-FEB-2002; 2002US-0357600P.
XX (EXEL-) EXELIXIS INC.
XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
PI Lioubin MN;
XX WPI; 2003-156840/15.
XX Identifying a candidate p53 pathway-modulating agent as therapeutic
PT targets for disorders related to defective p53 function e.g. cancer by
PT contacting an assay system having purified CHD polypeptide or nucleic
PT acid, with a test agent.
XX Claim 13; Page 229-237; 278pp; English.
XX The present invention relates to a method for identifying candidate p53
CC pathway modulating agents. The method involves contacting an assay system
CC comprising purified chromatin organisation modifier (chromo) domain
CC helicase DNA binding proteins (CHD), nucleic acids, their functionally
CC active fragments or derivatives, with a test agent under conditions
CC where, but for the presence of the test agent, the system provides a
CC reference activity. The methods are useful for identifying modulators of

CC the p53 pathway as therapeutic targets for disorders associated with
CC defective p53 function, such as angiogenic disorders, apoptotic disorders
CC or cell proliferative disorders, e.g. cancer. The modulators are useful
CC as research reagents, diagnostics and therapeutics. The invention is also
CC useful in gene therapy. The present sequence is human chromodomain
CC helicase DNA binding protein
XX
SQ Sequence 1912 AA;
Query Match 29.1%; Score 2091; DB 6; Length 1912;
Best Local Similarity 35.5%; Pred. No. 4.3e-149;
Matches 520; Conservative 205; Mismatches 424; Indels 314; Gaps 44;
Qy 21 DSDDDDDFVPKDRTPEQVEAIVRTDAKENACACGESNLVSCNTCTVAFHAKCLVPLP 80
Db 426 DNSEGEIEEIVGGDLEED-----DHHMEFCRVCKDGGELCCDTCPSGYHCHLNPPL 480
Qy 81 KDAVENWRCPECVSL--NEIDKILDCB--MRPTKSSSEQSGSDAE-----PKPIF--V 128
Db 481 PEIPNGEWLCPRCTCPALGKGVQKILIKWQGPSPPTVPRPPDADNPSPPLGRPE 540
Qy 129 KQYLVKWKGLSYLHCSVPPEK-----FQKAYKSNHRLKTRVNFHRQME----- 173
Db 541 RQFFVKWQGSYWHCSMVSELQLELHCQVMFRNYQRKNDMDPEPPSGDFGDEKSRKKN 600
Qy 174 ----SPNSEDDEP--VAIRPEWTVDRIL--ACREEDGELEYLVKYKELSYDECYWESE--- 224
Db 601 KDPKFAEMERFVRYGKPEWMMIHLNHSVDKKGHVHYLIKWRDLDPYDQASWESDEVE 660
Qy 225 -SDISTFONEIQRFKDV-----NSRTRRSKDVHDKENR-----DFQOFDHTPEFLK 270
Db 661 IQDYDLFKOSYNNHRELMEGEGRPCKKUKVKLRLEPPETPTVDTVTKYGRQREYLD 720
Qy 271 ---GLLHPYQLEGLNFRFSWSKQTHVILADEMGLGKTIQSIALLASLPEE--NLIPHLV 325
Db 721 ATGGTLHPYQMEGLNWLRFESWAGGTDITLADENGLGKTVQTAFLVLSYKEGHSKGPFLV 780
Qy 326 IAPLSLRLNWEREFATWAPQMVVMTFGTAQAPAVIREHFYLSKQOKIKKKKSGQISS 385
Db 781 SAPLSTIINWEREFEMWAPDMYVTVYVGDKSRATIRENEFEFEDNAIRGGKASRM--- 837
Qy 386 ESKQKEIKFDVLLTSYEMINLDSAVLKPIKWCMIVDGHRKLVKNDKSLPSSITQVSSNH 445
Db 838 -KKEASVKFHLVLTSYELTIDMAILGSDWACLIVDEAHLRLKNNQSKFPRVLNGSLQH 896
Qy 446 RILLTGTPLQNNLDELFLMHFLDAGKFGSLBEFQBEFKDINQEEQISRLHKMLAPHLR 505
Db 897 KULLTGTPLQNNLEELFHLNFLTPTFRHNLGFLBEFADIAKEDQIKKLHMLGPHMLR 956
Qy 506 RVKQDVNDMPKPKELILRLVDLSLQKEYYKAIPTFNQVLTQKGGG--QISLNNIMMELR 564
Db 957 RLKADVFKNMPKTELIIVRVLSLPMQKYKYKILTRNFEALNARGGNGVSLNLYVMDLK 1016
Qy 565 KVCCHPYMLEGVEPVTHDANEAPKQ-----LLESCGKLQLLDKMVKKEQGHVRL 615
Db 1017 KCCNHPYLF-----PVA--AMEAPKPNMGVDSALIRASGKLLQLLQMLKNEGGHRLV 1070
Qy 616 IYTFQFQMLDLLEDYCTHKKQYVERIDGKVGGAERQIRIDRFNNAKNSKFCFLSTRAGG 675
Db 1071 IFSQMTKMLDLEDLFLEHSGYKYVERIDGIGTGNMROEADIRFNAPGAQPCFLLSTRAGG 1130
Qy 676 LGINLATADTVIYDSNDWNPHADLQAMRAHRLGQNKVMYVRLINRGITIEERMQLTKK 735
Db 1131 LGINLATADTVIYDSNDWNPFHNDIQAFSRAHRTGQNKVMYVRFVTRASVEERITQVAKK 1190
Qy 736 KMWLEHLVVGK--LKTQNIQEEELDDIIRYGSKELFASDEDEAGSKG-----IHYD 785
Db 1191 KMWLTHLVYRPGLSGTSMKQELDDILKFGTEFLFKDEATDGGGDNKEGEDSSVIHYD 1250
Qy 786 DAAIDKLLDRDLVEAREVSDDEENGFLKAFKAVNFYIDENEAALBAQRVAASKSS 845
Db 1251 DXAIELLRDNDQETEDTEL--QGMNEYLSSEFKVQYVVVREEMGEVEEREIIRQEBS 1308

Qy	846	AGNSDRASYWEELLKKQFELHQAEELNALGKRRKRKQL-----VSIEDDLAAGLEDVSSD	901
Db	1309	VD-----PDYWEKLLRHHYEQQEDLARNLGGKRIKRVQVNNQSGSQDRD---WQDDQSD	1361
Qy	902	GDSESYAESIDTG--EAAGQGVOTGRPPYR---RKGRONLEPTPLMEGEGRSFVLGNQSQ	957
Db	1362	NQSDYSVASSEGDEDFDERSEAPRRPSRKGLRNDKKPLP--PLLARVGGNIEVLGFNARQ	1420
Qy	958	RAIFVOTLMRYG-----AGNFDKWEFVPRLKQKTFEINEYGLFLKHAIEEIDENSPTF	1012
Db	1421	RKAFLNAIMRYGMPQDAFTQW--LVYDLRGKSEKEFKAYVSLFMRHLCEPGADGAETF	1478
Qy	1013	SDGVPEKGLRIEDVLVRIALLILVQEKVKFVEDHPGKVPFPSRILFRPGLRSGKIWKEE	1072
Db	1479	ADGVPREGLSRQHLVTRIGVMSLIRKKVQEF-----	1510
Qy	1073	HDKIMIPAVLKHGYGRWQAIVDDKELGIGIELI CKELNFPHISLSAAFOAGLQONGSGGS	1132
Db	1511	-----HVNGRW-----SMP--ELAEBEENKKMSQPGS--P	1536
Qy	1133	NPGAQTNONPGSVITGNNNASADGAQVNSMFYYRDMQRRLVEFVKRVLLEKAMNYEYA	1192
Db	1537	SPKTPPTSTFGD-----TQNT-----	1553
Qy	1193	EYYGLGGSSSIPTEEPAEPKPIADTVGVSFIEVDDMDLDPKTDPTITSEINGAAVDN	1252
Db	1554	-----PAPVPPAE-----DGIK-----	1565
Qy	1253	NQARVETAHYNQMKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNI-NTILSI	1311
Db	1566	-----IEN-----SLKEESIEGEKEVKSTAPETAIEC	1594
Qy	1312	TSQSKSHEDDT-----KPDLLNNVEMKDTAE---ETKPLRGVVVDLNNVGEENIAE	1360
Db	1595	TQAPAPASEDEKVVVEPPEGEKEVKEAEVKEETEEPMETEP--KGAADVEKVE-EKSAID	1651
Qy	1361	ASGSVDVVMEEAEKEEEKPKQMVV	1383
Db	1652	LTPIVVEDKEEKEEKEEKEVWL	1674
RESULT	5		
AAE36109	ID	AAE36109 standard; protein; 1912 AA.	
XX	AC	AAE36109;	
XX	DT	26-JUN-2003 (first entry)	
XX	DE	Human chromodomain helicase DNA binding protein (CHD) #7.	
XX	KW	Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;	
XX	KW	chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;	
XX	KW	cell proliferative disorder; chromatin organisation modifier domain;	
XX	OS	cytostatic.	
XX	OS	Homo sapiens.	
XX	PN	W0200298999-A2.	
XX	PD	12-DEC-2002.	
XX	PF	03-JUN-2002; 2002WO-US017466.	
XX	PR	05-JUN-2001; 2001US-0296076P.	
XX	PR	10-OCT-2001; 2001US-0328605P.	
XX	PR	22-OCT-2001; 2001US-0338733P.	
XX	PR	15-FEB-2002; 2002US-0357253P.	
XX	PR	15-FEB-2002; 2002US-0357600P.	
XX	PA	(EXEL-) EXELIXIS INC.	
XX	PI	Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;	

```
Db 1017 KCCNHPYLF---PVA--AMEAPKPMNGMYDGSALIRASCKLLLLQKMLKNLKEGGHVL 1070
Qy 616 IYTOFQMLLLEDDYCTHKWQYERIDGKVGGAERQIRIDRFNAKNSNKFCLLSTRAGG 675
Db 1071 IFSQMTKMLDLEDFLEHEGYKYERIDGGITGNMROEAIIDRFNAPCAQQCFLLSTRAGG 1130
Qy 676 LGINLATADTVIYDSDWPHADLOAMARAHRLGQTNKVMYIRLINRGTIERERMOLTKK 735
Db 1131 LGINLATADTVIYDSDWPHNDIQAFSRAHRIQONKVMYIRFVTRASVEERITQVAKK 1190
Qy 736 KMWLEHLVVGK---LKTQNTQNOBELDDIIRYSGKELFASDEDDBAGSGK-----IHYD 785
Db 1191 KMWLTHLVVRPGLSGTKGNSKQELDDILKFGTEELFKDEATDGGDNKEGEDSSVIHYD 1250
Qy 786 DAIDKLLDRDLVAEBEVSDDDEENGFLKAFKVANFEYIDENEAALAEQORVAASKSS 845
Db 1251 DKAIERLLDRNQDETEDEL--QCMNEYLSFFKAYQVVRREEMGBEEVEREIIKQES 1308
Qy 846 AGNSDRASYWEELLKOKFELHQAELNALGKRKRKQL---VSTEEDDLAGLEVDSSD 901
Db 1309 VD-----PDYWEKULRHHYEQOQEDLARNLKGKRIKQVNYNDGSGQEDRD---WQDDQSD 1361
Qy 902 GDSYEAESTDG--EAAGQGVQGTGRRPYR---RKGRDNLEPTPLMEGEGRSFRVLGFNQSQ 957
Db 1362 NQSDYSVASSEGDEDFDERSEAPRPSRKGRLNDKOKPLP--PLLARVGNIIEVLGFNARQ 1420
Qy 958 RAIFVQTLMYRG-----AGNFDWKEFVRLKQKTFEINIEYGLFLKHAEEIDENSPTF 1012
Db 1421 RKAFLNAIMRYGMPPODAFTQW--LVRDLRGKSEKEFKAYVSLFWRHLCEPGADGAETF 1478
Qy 1013 SDGVPKEGLRIEDVIRIALLILVQEKVFEVDHPGKFPVPSILERFPFLRSGKTKWEE 1072
Db 1479 ADGVPREGLSRQHVLRIGVMSLIRKKVQEF----- 1510
Qy 1073 HDKIMIRAVLKHGVRQWQAIIVDDKELGICQLI CKELNFPHISLSAAEOAGLOGQSGGS 1132
Db 1511 -----HNGRW-----SMP--ELAEVEENKKMSQPGS--P 1536
Qy 1133 NPGAQTNQNPQSVITGNNNASADGAQVNSMFIYDMORRLVEFVKRVLLEKAMNYEA 1192
Db 1537 SPKTPPTSTPGD-----TQDNT----- 1553
Qy 1193 EYYGLGSSSITTEPEAPKIAIDTVGVSFIEVDDMDLGLPKTDPITSEETMGAAVDN 1252
Db 1554 -----PAPVPAP-----DGIK----- 1565
Qy 1253 NQARVEIAQHYNQCKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNI--NTILSI 1311
Db 1566 -----IEEN-----SLKEESEIEGEKEVKSTAPETAIEC 1594
Qy 1312 TSDQSKSHEDDT-----KPLNNVEMKDTAB---ETKPLRGVVVDLNVVEGEENIAE 1360
Db 1595 TQAPAPASEDEKVVVEPEEGEKEVKAERVTEEPMETEP--KGAADVKEVB--EKSAID 1651
Qy 1361 ASGSVDVWKEEAKKEEKPMWV 1383
Db 1652 LTPIVVEDKEEKEEKEKGVML 1674

RESULT 6
AAR99534
ID AAR99534 standard; protein; 1911 AA.
XX
AC AAR99534;
XX
DT 30-OCT-1996 (first entry)
XX
DE Dermatomyositis specific autoantigen, Mi-2.
XX
KW Mi-2; autoantigen; collagen disease; chromosome 12; 12p13; helicase;
KW dermatomyositis; diagnosis.
XX
OS Homo sapiens.
```

```
XX Key Location/Qualifiers
FH Region /label= Region_a
FT /note= "contains 3 potential core target motifs"
FT 113. .133
FT /label= Region_b
FT /note= "contains 4 potential core target motifs"
FT 133. .143
FT /note= "possible electrostatic interaction with chromatin
of histones"
FT 257. .287
FT /label= Region_c
FT /note= "contains 3 potential core target motifs"
FT 747. .758
FT /note= "Helicase-specific motif I"
FT 782. .793
FT /note= "Helicase-specific motif IA"
FT 869. .877
FT /note= "Helicase-specific motif II"
FT 897. .911
FT /note= "Helicase-specific motif III"
FT 943. .959
FT /label= Region_d
FT /note= "contains 1 potential core target motif"
FT 949. .960
FT /note= "Helicase-specific motif IV"
FT 1121. .1144
FT /note= "Helicase-specific motif V"
FT 1149. .1183
FT /note= "Helicase-specific motif VI"
XX DE19509279-Cl.
PN 15-MAY-1996.
XX 15-MAR-1995; 95DE-01009279.
PF 15-MAR-1995; 95DE-01009279.
PR 15-MAR-1995; 95DE-01009279.
XX (PRIV-) PRIVATES INST IMMUNOLOGIE & MOLEKULARGEN.
XX Seelig HP, Renz M;
XX WPI; 1996-240280/25.
DR N-PSDB; AAT32301.
XX DNA encoding dermatomyositis specific auto:antigen - useful for
differential diagnosis and treatment of dermatomyositis.
XX Claim 1; Fig 2; 20pp; German.
XX The present sequence is that of a 218 kD dermatomyositis specific auto-
antigen, designated Mi-2. The sequence numbering given in the
specification starts at amino acid 2, i.e. the first Met residue is
omitted. The protein is hydrophilic, acidic and protruding regions of the
protein are characteristic of helicases. The gene corresponding to the
cDNA (AAT32301) encoding Mi-2, was localised to chromosome 12 (12p13).
XX The DNA can be used for the recombinant production of Mi-2 which is used
for, e.g. the differential diagnosis of collagen diseases, esp.
dermatomyositis, e.g. by immunoassay or Western blotting
XX Sequence 1911 AA;
```

```
Query Match 29.1%; Score 2088; DB 2; Length 1911;
Best Local Similarity 35.5%; Pred. No. 7.3e-149;
Matches 519; Conservative 206; Mismatches 424; Indels 314; Gaps 44;
Qy 21 DSDDDDDFPVKRDTPEQVEAIVRTDAKENACQAGCESTNLVSCNTCTYAFHAKCLVPL 80
Db 425 DNSEGEIELEVGGDLEED----DHMEFCRVCKDGGELCCDCTCPSSYHIHCLNPPL 479
Qy 81 KQASVENWRCPCVSP---NEIDKILDCE--MRPTYKSEQSSDAB-----PKPIF---V 128
```


Db 480 PEIPNGEWLCPRTCTPALKGVKQKILITWKWQPPSPTPVPRPPDADNTSPKPLEGRPE 539
Qy 129 KOYLKWKGLSYLHCSWVPEKE-----FQKAYKSNHLKTRVNNHFQME----- 173
Db 540 RQFVKQWGSYWHCSWSLQELHCVQNFVYQKNDWDEPPSGDFGGDEBKSRKKN 599
Qy 174 ---SFNNSDDF--VAIRPWTTVDRIL-ACREDEGELEYLVKYKELSYDECYWESF--- 224
Db 600 KDPKFAEMEERFVRYGIGKPEWMIHRLNHSVDKGVHVLIKWRDLPYDQASWESDEVE 659
Qy 225 -SDISTFQNIQPKOV-----NSRTRRSXNDVDHKNR-----DFOQPDHTEFLK 270
Db 660 IQDYDLFKQSYWNRHRELMRGEGRPKKLKKVLRKLERPPETPTVDPVTKYERQPEYLD 719
Qy 271 ---CLHPYQLEGNFLRFSSKQTHVLADENGLKTIQSIALLASLPEE--NLPHLV 325
Db 720 ATGTLHPYQWGLNWLRFSAQGTDTLADENGLKTIQVAVFLSYLKEGHSKGFLV 779
Qy 326 IAPLSTLRNWEREFATWAPQNVVMYFGTAQARAVIREHEFYLSKQKKIKKKSGQISS 385
Db 780 SAPLSTIINWEREFEMWAPDMYVTVYGDKDSRAIIRENEFSFEDNAIRGGKASRM--- 836
Qy 386 ESKOKRIKFDVLTYSYEMINLDSAVLKPIKWEKMIVDEGHLKXNOSKLFSSLTQYSSNH 445
Db 837 -KKEASVKFVLLTSYELITIDMAILGSIWACLIVDEAHLKXNOSKLFVRVNGYSLOH 895
Qy 446 RIILITGTPLONLDELPMWHFLDAGKFGSLEFQBEFKDINQEBEQISRLHKLAPHILR 505
Db 896 KLLITGTPLONLDELPHLLNFTPERFNLLEGFLFEEFADIAEDQIKLHDLMPHMLR 955
Qy 506 RVKDVNMKMPKELLIRLDLSLOKEYYKAIPTFRNYQVLTKKGA-QISLNNIMMELR 564
Db 956 RLKADVFNKPSKTELIVRVELSPMQKYYKILTRNFEALNARGGNQVSLNVMDLK 1015
Qy 565 KVCCHPMLGSEVPVHIDANEAFKQ-----LLESCGKQLLDMMVKLKEQGHVRL 615
Db 1016 KCCNHPYLF---PVA--AMEAPKPMGMYDGSALIRASGKLLQLKMLKXKEGGHVL 1069
Qy 616 IYTOFQHMULDLEDYCTHKWQVERIDGKVGGAEROIRIDRENAKNSKKEFCILLSFRAGG 675
Db 1070 IFSQMTKMLDLEDFLEHEGYKERYIDGGITGNRQEAIRFNAPGAQQFCFLLSFRAGG 1129
Qy 676 LGINLATADTVIYDSWNPHADLOAMARAHRLQTNKVMIRYLINRTGIEERMMLTKK 735
Db 1130 LGINLATADTVIYDSWNPHNDIQAFSAHRIGQNKVMIRYVTRASVEERITQVAK 1189
Qy 736 KMLVLEHLVVGK---LKTQINQBELDDIIRYGSKELFASDEDEBAGSKG-----IHYD 785
Db 1190 KMLVLEHLVVRPGLSGTKGSMKQELDDILKFGTEELFKDEATDGGGDNKEGESSVHYD 1249
Qy 786 DAADKLLDRDLVAEEVSDDEENGFLAKFANFEYIDENEAALAEQORVAESKSS 845
Db 1250 DKAERLLDRNQDETEDEL--QGMNEYLSFSFKVAQYVVRREEEMGEEDVEREIIKQEE 1307
Qy 846 AGNSDRASYWEELKDKFELHQAELNALGKRKRKQL---VSIREDDLALEDVSSD 901
Db 1308 VD----PDYNEKLRLHHYEQOEDLARNLKGKRIKQVNYNDGSQEDRD---WQDDQSD 1360
Qy 902 GDESYEABSTG--BAAGQGVQVGTGRPVYR---RKGRDNLPTPLMEGEGRSFVRLGPNQSQ 957
Db 1361 NQSDYSVASBEGEDDFERSEAPRPSRKLGRNDKDKPLP--PLLARVCGNIEVLGPNARQ 1419
Qy 958 RAIFVQTLMYG-----AGNFDWKEFVPLKQKTFEINEYIGILFLAHIAEEDIDENSPTF 1012
Db 1420 RKAFNLAIMRYGMPPODFTQW--LVRDLRGSEKEFKAYVSLFMRHLCEPQADGAET 1477
Qy 1013 SDGVPKEGLRIEDVLRIALLILVQEKVFEVDHPGKVPFPPSRILERFPGRLSGKIWKEE 1072
Db 1478 ADGVPREGLSHQHVLTRIGVMSLRKVKQBEF----- 1509
Qy 1073 HDKIMIRAVLKHGVRQWQIVDDKELGIGQELICKELNFPNPHISLSAAEQAGLQGNQSGGS 1132

Db 1510 -----HVNGRW-----SMP--ELAAVEENKMSQPSG--P 1535
Qy 1133 NPGAQTNQPGSVITGNNASADGAQVNSMFWYVYRDQRRLLVEFKRVLILLEKAMNYEA 1192
Db 1536 SPKTPPTSPTEGD-----TQPN----- 1552
Qy 1193 EEEYGLGSSSTPTEPEAPKTIADTVGVSVFIEVDDMLDGLPKTDPITSEEMGAVDN 1552
Db 1553 -----PAPVPPAE-----DGIK----- 1564
Qy 1253 NQARVEIAOHYNQCKLLENARESVQAVYNNQPPSTKVNESPRALKSINGNI--NITLSI 1311
Db 1565 -----IEN-----SLKEEESTEGSKVKSTAPETAIEC 1593
Qy 1312 TSDQSKSHEDDT-----KPDLNVMEMKDTAE---ETKPLRGVVDLNVVGEENIAE 1360
Db 1594 TQAPAPASDEKVVFPPEGEKVEKAEKVERTSEEMETEP--KGADVEKVE--EKSAID 1650
Qy 1361 AGSVDVDMKEEAKKEEKPKNMVV 1383
Db 1651 LTPVVVDKEEKEEKEKEVML 1673
RESULT 7
AAM39288
ID AAM39288 standard; protein; 1912 AA.
XX AAM39288;
AC AAM39288;
XX 22-OCT-2001 (first entry)
DT Human polypeptide SEQ ID NO 2433.
XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX Homo sapiens.
OS WO200153312-A1.
XX 26-JUL-2001.
XX 26-DEC-2000; 2000WO-US034263.
XX 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR N-PSDB; AAI58444.
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX Example 4; SEQ ID NO 2433; 10078pp; English.
PS The invention relates to human nucleic acids (AAI57798-AAI61369) and the
XX

Query Match	28.8%;	Score 2067.5;	DB 4;	Length 1936;
Best Local Similarity	35.3%;	Prod. No. 2.7e-147;		
Matches 517;	Conservative 205;	Mismatches 429;	Indels 313;	Gaps 44;
Qy	21	DDSDDDDFVFKDRTFQVRAIVRTDAKENACQAGBSTNLVSNVCTCTYAFHAKCLVPL 80		
Db	447	DNEGEIILSEVGDDLEED-----DHHMEFCRVCKDGGELLCDCTCFSSVHIICLAPPL 501		
Qy	81	KDASVENWRPECVSP--NEIDKILCE--MRPTKSEQSSDAE-----PKPIF--V 128		
Db	502	PEIPNGEWLCPRCTCPALKGVOKILIKWGPSPPTVPVRPPDADNTPTSPKPLEGRPE 561		
Qy	129	KOYLVKWKGLSYLHCSVPKE-----FQAKYSNHLKTRVNNFHRQME----- 173		
Db	562	RQFVFKQWGSYHCHSWSVLSQLELHCQVNFNRYQRKNDMDDEPPSGDFGDEEKSRRKN 621		
Qy	174	--SPNNSDDF--VAIRPWTVDTRL-ACREBGELEYLVKYSLEYDCEYWESE--- 224		
Db	622	KDPKFAEMERFVRYGLKPMWMIHRLNHSVDKKGHHVYLIKWRDILPYDOASHESDEVE 684		
Qy	225	-SDISTFQNEIORFKV-----NSRTRRSKDVDHKNRPR-----DFOQFDHTPBFLK 270		
Db	682	IQDYDLFKQSYWNHRELMRGEGRPGKKLVKLRLPERPETPTVPTVKYERQPEYLD 741		
Qy	271	---GLLHPYQLEGLNFIKFSWSQTHVILADEWGLGKTIOSIALLASLFE--NLI 325		
Db	742	ATGCTLHPYQMEGLNMLRFSWAQOTDTILADEMGLGKTQTAPEVFLSYLYKEGHSKGPFLV 801		
Qy	326	IAPLSTLRNWEREFATWAPOMNVMYFGTAQAAVIREHEFYLSKQOKKIKKKSGQISS 385		
Db	802	SAPLSTLRNWEREFEMWAPDMYVTVYGDKDSRAIIRENEFSFEDNAIRGKKASRM--- 858		
Qy	386	ESKQRIKIFDVLTSYEMINLDSAVLKPIKWCMIVDEGHLKNKDSKLFSSLTQYSSNH 445		
Db	859	-KKEASVKFHVLTSEYLIITIDMAILGSDIWACLIVDEAHLKNNQSKFPRVLNGYSIQH 917		
Qy	446	RIILTGTPLONNDELFLMLMHFDAGKFGSLEBFQBEFKDINOBEQISRLHKMLAPHLLR 505		
Db	918	KLLTGTPLQNNLEELFHLNFLTPTFRHNLLEGFLFEFADIADQDKLCLDMLGPHMLR 977		
Qy	506	RVKQVQWKMPPKELITRLVDLSLOEYKKAIFTRNY-QVLTKKGA-QISLNNIMMEL 563		
Db	978	RLKADVFNKMPSTELIVRVELSPQKKYKYILHSLFKALNARGGNQVSLNVVMDL 1037		
Qy	564	RKVCCHPYMLEGVEPVITHDANEAFKQ-----LLESCGQLQLLDKMMVKLKEQGHV 614		
Db	1038	KKCCNHPYLF---PVA--AWEAPKPNMGVYDGSALIRASGKLLLLQKMLKNLKEGHRV 1091		
Qy	615	LIYTFQOHLMDLLEDYCTHKWQYERIDGKVGAEQRIIRDFAKNKSNKPCFLLSTRAQ 674		
Db	1092	LIFSQMTKMLDLEDFLEHEGYKYERIDGGITGMROEADIRFNAPGAQFCFLLSTRAQ 1151		
Qy	675	GLGINLATDVIILYDSDNPHADLQMAAHLRGQNKVMIYRLINRGITIEERMQLTK 734		
Db	1152	GLGINLATDVIILYDSDNPHNDIQAFSAHRIGQNKVMIYRVFVTRASVEERITQAK 1211		
Qy	735	KMWLEHLVVGK--LKTQINOBELDDIIRYGSKELFASDEDAKSGK-----IHY 784		
Db	1212	KQMLTHLVVRPGLSGTGSMSKQELDDILKFGTEELFKDEATDGGDNKEGEDSSVHY 1271		
Qy	785	DDAIDKLLDRDILVBAEVSVDDEENGFLKAFKVFANFEYIDENEAALAEQRAVAESKS 844		
Db	1272	DDKAIERLLDRNQDETDTEL--QGMNEYLSSFKVAQVYVREEMGEVEEREIIKQEE 1329		
Qy	845	SAGNSDRASVWELLKDKFLHQABELNALGKRKRKQI-----VSTEEDDLAGLEDVSS 900		
Db	1330	SVD-----PDYWEKLLRHYYEQQEDLARNJGKGRIRKQVNYNDGSOEDRD---WQDDQS 1382		
Qy	901	DGDESVEAESTDG-EAAGQGVQTRPYR---RKGRDNLSEPTPLMEGEGRSFVLVGNQS 956		
Db	1383	DNQSDVSVASEGDEDFDERSEAPRPSRKLNRNDKDKPLP-PULARVGNIEVLGNAR 1441		
Qy	957	QRAIFVQTLNRYG-----AGNFQWKEFVPRLKQTKTEINEYGILLFKHTAEEIDENSPT 1011		

QY	1286	PP-STKVNESFRALKS	1300
	: : :		
Db	1856	PPIAARLQMSERSILS	1871
RESULT 9			
AAW41074			
ID	AAW41074	standard; protein; 1936 AA.	
XX	AC	AAW41074;	
XX	DT	22-OCT-2001 (first entry)	
XX	DE	Human polypeptide SEQ ID NO 6005.	
XX	DE		
XX	KW	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;	
XX	KW	peripheral nervous system; neuropathy; central nervous system; CNS;	
XX	KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
XX	KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
XX	KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;	
XX	KW	leukaemia.	
OS	OS	Homo sapiens.	
PN	PN	WO200153312-A1.	
XX	XX	26-JUL-2001.	
XX	XX	26-DEC-2000; 2000WO-US034263.	
XX	XX	23-DEC-1999; 99US-00471275.	
PR	PR	21-JAN-2000; 2000US-00488725.	
PR	PR	25-APR-2000; 2000US-00552317.	
PR	PR	20-JUN-2000; 2000US-00598042.	
PR	PR	19-JUL-2000; 2000US-00620312.	
PR	PR	03-AUG-2000; 2000US-00653450.	
PR	PR	14-SEP-2000; 2000US-00662191.	
PR	PR	19-OCT-2000; 2000US-00693036.	
PR	PR	29-NOV-2000; 2000US-00727344.	
PA	(HYSE-)	HYSEQ INC.	
XX	XX		
PI	PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
PI	PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;	
PI	PI	Zhou P, Goodrich R, Drmanac RT;	
XX	XX	WPI; 2001-442253/47.	
XX	XX	N-PSDB; AAI60230.	
PT	PT	Novel nucleic acids and polypeptides, useful for treating disorders such	
PT	PT	as central nervous system injuries.	
XX	XX	Example 2; SEQ ID NO 6005; 10078pp; English.	
PPS	PPS	The invention relates to human nucleic acids (AAI57798-AAI61369) and the	
XX	XX	encoded polypeptides (AAW38642-AAW42213) with nootropic.	
CC	CC	immunosuppressant and cytostatic activity. The polynucleotides are useful	
CC	CC	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	CC	of the invention may be used to treat diseases of the peripheral nervous	
CC	CC	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	CC	localised neuropathies and central nervous system diseases, such as	
CC	CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	CC	utilisation of the activities such as: immune system suppression,	
CC	CC	Activ/inhibin activity, chemotactic/chemokinetic activity, haemostatic	
CC	CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
CC	CC	assays for receptor activity, arthritis and inflammation, leukaemias and	
CC	CC	C.N.S disorders. Note: The sequence data for this patent did not form	
CC	CC	part of the printed specification	
XX	XX	Sequence 1936 AA;	
XX	XX	Sequence 1936 AA;	

Db 1442 QRAFLNAINRYGMPQDAFTQW--LVDRLGKSEKEFRAYVSLFMRHLCEPCADGAET 1499
 Qy 1012 FSDGPKGELRIEDLVRIALLILVQEKVKFVEDHPGKVPFPSPILLERFPGLRSGKIWE 1071
 Db 1500 FADGVPREGLSRQHLVLRIGVSLIRKKVOEFE----- 1532
 Qy 1072 EHDKIMIRAVLKGYGGRQAIQVDDKELGIGELICKELNPFPHISLSAAEQAGLQGGSGG 1131
 Db 1533 -----HVNGEW-----SMP--ELAEVENKMSQPGS-- 1557
 Qy 1132 SNPGAQNNQNGSVITGNNNASADGAQVNSFYRDQMRLVBFVKRVLLEKAMNVEY 1191
 Db 1558 PSPKTPPTSPGD-----TQPN----- 1575
 Qy 1192 AEEYGLGGSSSTPEPEAPKPIADTVGVSFIEVDMDGLPKPTDPTISBEIMGAAVD 1251
 Db 1576 -----PAPVPAE-----DGK----- 1587
 Qy 1252 NNQARVEIAQHYNQMKLLDENARESVOAYVNNOPSTKYNESFRALKSINGNI-NTILS 1310
 Db 1588 -----ISEN-----SLKEESIEGEKVKSTAPETAIE 1615
 Qy 1311 ITSDQSKSHEDT-----KPLNNVEMKOTAE-----ETKPLRGVVDLNVVEGENIA 1359
 Db 1616 CTQAPAPASDEKVVVBPPEGEKVEKAEKTERTEPEMETEPKKGGAADVKEVE-EKSAI 1674
 Qy 1360 EASGSVDVKEEAKKEEKPKMNV 1383
 Db 1675 DLTPVIVVEDKEEKEEKEVML 1698

RESULT 10

ABP64883
 ID ABP64883 standard; protein; 1470 AA.

XX AC ABP64883;
 DT 25-FEB-2003 (first entry)
 DE Human protein SEQ ID 543.
 XX Human; expressed sequence tag; EST; haematopoietic disorder;
 KW central nervous system disease; viral infection;
 KW peripheral nervous system disease; non-healing wound; infectious disease;
 KW immune deficiency; immune disorder; bacterial infection; allergy; cancer;
 KW fungal infection; autoimmune disorder; coagulation disorder; neutropenic;
 KW antiallergic; antiinflammatory; immunosuppressive; neuroprotective;
 KW cytostatic; haemostatic; virucide; antibacterial; fungicide;
 KW immunostimulant; cerebroprotective.
 XX Homo sapiens.
 XX WO200259260-A2.
 XX 01-AUG-2002.
 XX 16-NOV-2001; 2001WO-US042950.
 XX 17-NOV-2000; 2000US-00714936.
 XX (HYSE-) HYSEQ INC.
 PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA;
 PI Ren F, Xue AJ, Yang Y, Wehrman T, Drmanac RT;
 XX WPI; 2002-590824/63.
 DR N-PSDB; ABQ99469.
 XX New isolated polynucleotide, useful in research, diagnostic or
 PT therapeutic methods, e.g. preventing or treating disorders involving
 PT aberrant protein expression or biological activity.

Claim 20; SEQ ID NO 543; 394pp; English.

The present invention relates to novel human coding sequences (ABQ99268-ABQ99608) and proteins (ABP64682-ABP65022). The sequences are useful in therapeutic, diagnostic and research methods. The polynucleotides may be used in the field of molecular biology as hybridisation probes, primers for PCR, for chromosome and gene mapping, for the recombinant production of protein, or in generation of anti-sense DNA or RNA. The polynucleotides are useful in diagnostics as expressed sequence tags (ESTs) for identifying expressed genes or for physical mapping of the human genome. The proteins may be used as molecular weight markers, or as nutritional sources or supplements. The proteins may be used to maintain and expand cell population in a totipotent or pluripotent state useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals or the development of bio-sensors. The polynucleotides and proteins are useful for preventing, treating or ameliorating disorders involving aberrant protein expression or biological activity, e.g. haematopoietic disorders, central/peripheral nervous system diseases, mechanical and traumatic disorders, non-healing wounds, immune deficiencies and disorders, infectious diseases caused by viral, bacterial or fungal infection, autoimmune disorders, allergic reactions and conditions, coagulation disorders, or cancer. The polynucleotide sequences of the invention were assembled from ESTs isolated mainly by sequencing by hybridisation, and in some cases, sequences obtained from one or more public databases. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 1470 AA;

Query Match 27.9%; Score 2007; DB 5; Length 1470;

Best Local Similarity 36.6%; Pred. No. 7.1e-143;

Matches 503; Conservative 212; Mismatches 387; Indels 272; Gaps 43;

Qy 123 PKP---IFVKQYLKVKGLSYLHCSWVPEKFKAYKSNHRLKTRVNFHR----- 170
 Db 16 PKPLEGIPEREFFVKWAGLSYHCSWVKELQLELYHTVMYRQKNDMDPEPPFYDGS 75
 Qy 171 ----QWESFNNS-----EDDF--VAIRPWTTVDRIL--ACREDEGELEYLVKYSLS 216
 Db 76 DEGKSEKKNKDPDYAKMEERYFYGIKPEWMMHRLNHSFDKGDVHYLLKWKDLPY 135
 Qy 217 DECYWE--SESDISTFQNEIQRF-----KOVNSRTRSKVDVHKRNPDRFQ 260
 Db 136 DQCTWEIDDDIDIPYDNLQAYWGHRELMLGDTLPLKRLKKGKLLRD--DKQEKPDTP 194
 Qy 261 QFHTPEFLK-----GLLHPYQLEGFLNFRFSKQTHVLADMGKTIQSIAL 311
 Db 195 IVDPTVKFDKQPYIDSTGTLHPYQLEGFLNFRFSWAQGTDTILADEMGLKTVQTVIF 254
 Qy 312 LASLFEF--NLIPHVIAPLSTLRNWEREFATWAPQWVNYFGTAQARAVIREHEFYS 369
 Db 255 LYSLYKEGHSKGPILVSAPLSTIINWEREFEMWAPDFYVVTYTGDKESRVIRENEP--S 312
 Qy 370 KDQKIKKKKSGQISSESKOKRIKFDVLLTSYEMINLDSAVLKPIKWECMIVDEGHLKN 429
 Db 313 FEDNAIRSGK--KVFRMKKEVQIKFHVLLTSYELLITDQAILGSIWACLWVDEAHLKN 370
 Qy 430 KSKLPSLLTOYSSNHRILLTGTPLNLDLFLMHLFDAGKFSLEEFQEFKIDNOE 489
 Db 371 NQSKFRVLNSYKIDYKLLTGTPLNLDLFLMHLFDAGKFSLEEFQEFKIDNOE 430
 Qy 490 EQISRLHKLAPHLRRVKDKVMKMPKPKELILRVLDLSLQKEYYKAIETRYVQVLTTK 549
 Db 431 DQIKLHDLGLGPHMLRLKADVFKNPAKTELIVRVELSQKYYKFIITRFEALNSK 490
 Qy 550 GGA-QISLNNIMMELRKVCCHPYM--LEGVE--PVIHDANFAKQLLESCKQLQLLDRQMV 605
 Db 491 GGGNQVSLNIMMDLKKCCNHPYLFVAAVEAPVLPNGSYDGSLSVSKGKMLLQKMLK 550
 Qy 606 KLKEQHRVLITYTQFQHMLDLLEDYCTHKKQWYRIDGKVGGAERQIRIDRFKNSKNKF 665

```
Db 551 KLRDEGHRVLI FQMTQKMLDLEDFLEYEGYKVERIDGGITGGLRQEAIDRFNAPGAQQF 610
Qy 666 CFIILSTRAGGLGINLATADTVIIYDSWNPADLOAWARHRLGQTNKWMYILNRGTI 725
Db 611 CFIILSTRAGGLGINLATADTVIIYDSWNPADLOAWARHRLGQTNKWMYILNRGTI 670
Qy 726 EERMWOLTKKKWVLEHLVCK--LKTONTNOBELDDIIRYSGKELFAS-----772
Db 671 EERITQVAKRKMMLTHLVVRPGLSGSGMTKQELDILKFGTEELFKDQVEGMSQGOR 730
Qy 773 -----DDEAGKSGK-----IHYDDAAIDKLLDRDLVEAE 801
Db 731 PVTPIDVQSSKGNLAASAKKHGTPPGDNKDVEDSSVIHYDDAAISKLLDRN-----795
Qy 802 EVSVDDE-----ENGFLKAFKANFEYIDENEAALQRAVAESKSGASGNSDRASWHEEL 858
Db 786 QDATDDTELONNNEYLSFSAQYVQVREEDGVEVEREIKQEE-----NVD-PDYWEKL 839
Qy 859 LKDKFELHQAELNALGKRKRKQL-----VSTEEDDLACLEDVSDGDESY-----RAES 910
Db 840 LRHHYEQQBDLARNLGKRIKQVNYNDASQDQE-----WQELSDNQSEYISGEDEED 896
Qy 911 TDGEAAGQGVQTRRPRRK---GRNLEPTPLMEGEGRSFRVLGFGNQSORAIFVQTLMR 967
Db 897 EDFEERPEG-QSGRRQSRRLKSDRKPLP-PLLARVGGNIEVLGFNARQKAFNAIR 954
Qy 968 YG-----AGNFDKWEFVRLKQKTFREINBYGILFKHAEIDENSPTSDGVKPEGLR 1022
Db 955 WGMPPQDAFNHW--LVRLDRGKSEKEFRAYVSLFMRHLCEPGADGAETPADGVPREGLS 1012
Qy 1023 IEDVLVRIALLIIVQEKVKEVDHPKVPFPPSILRFRFGLRSGKIWEHDKIMIRAVL 1082
Db 1013 RQHLVTRIGVMSLVKRVKQVEFHVNGKYSTPDLPPEGPKKPGEV-----1058
Qy 1083 KHGYGRWQAIVDDKELGI-----QELICKELNFPHISLSAAEQAGLOQNGSGSNPQAOT 1138
Db 1059 -----ISSDPNTPVPASPAPHLPAPLGPD---KWEAQLGYMDE-----KDFGAQK 1101
Qy 1139 NQPGSVITGNNASADGAQVSNFYRDMQRRLVFVKRVLLBKAMNVEABEYGL 1198
Db 1102 PRP-----LEVOALPAALDRVESDKH--1124
Qy 1199 GSGSSIPTREPAEKIADTVGVSFIEVDMDLGLPKTPTITSEIMGAVDNQNQARVE 1258
Db 1125 -----ESPASKEARE-----ERPETEKAAPPSPQLPREVL-----1157
Qy 1259 IAQHYNQMKLLDENARESVOAYVNN---OPPSTKVNESFRALKSINGNINTILSITSDQ 1315
Db 1158 -----PEKEKILDKLSELHSGDSSELPPDDTKAEKEPIETQNGD-----1201
Qy 1316 SKSHEDDTKPLDNNVEMK-----DTAETKPLRGVVDLNVV-EGEENIABASGSV 1365
Db 1202 -KEEDDEGKKEDKKGKFKFNFNIAD-----GGFTLHTLWQNEERAAVSSGKI 1248

RESULT 11
ID AAE36108
AC XX
AC AAE36108;
XX
DT 26-JUN-2003 (first entry)
XX
DE XX
DE XX
KW Human chromodomain helicase DNA binding protein (CHD) #6.
KW Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;
KW chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;
KW cell proliferative disorder; chromatin organisation modifier domain;
KW cytoskeletal.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
```

```
FT Domain 12..54
FT /note= "Chromodomain"
FT Domain 126..422
FT /note= "SNF2N domain"
FT Domain 487..561
FT /note= "Conserved C-terminal domain"
XX
XX WO200298899-A2.
PN 12-DEC-2002.
XX
XX 03-JUN-2002; 2002WO-US017466.
XX
XX 05-JUN-2001; 2001US-0296076P.
PR 10-OCT-2001; 2001US-0328605P.
PR 22-OCT-2001; 2001US-0338733P.
PR 15-FEB-2002; 2002US-0357253P.
PR 15-FEB-2002; 2002US-0357600P.
XX
XX (EXEL-) EXELIXIS INC.
XX
XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
PI Lioubin MN;
XX
XX WPI; 2003-156840/15.
XX
XX Identifying a candidate p53 pathway-modulating agent as therapeutic
XX targets for disorders related to defective p53 function e.g. cancer by
XX contacting an assay system having purified CHD polypeptide or nucleic
XX acid, with a test agent.
XX
XX Claim 13; Page 215-221; 278pp; English.
XX
XX The present invention relates to a method for identifying candidate p53
XX pathway modulating agents. The method involves contacting an assay system
XX comprising purified chromatin organisation modifier (chromo) domain
XX helicase DNA binding proteins (CHD), nucleic acids, their functionally
XX active fragments or derivatives, with a test agent under conditions
XX where, but for the presence of the test agent, the system provides a
XX reference activity. The methods are useful for identifying modulators of
XX the p53 pathway as therapeutic targets for disorders associated with
XX defective p53 function, such as angiogenic disorders, apoptotic disorders
XX or cell proliferative disorders, e.g. cancer. The modulators are useful
XX as research reagents, diagnostics and therapeutics. The invention is also
XX useful in gene therapy. The present sequence is human chromodomain
XX helicase DNA binding protein
XX
XX Sequence 1377 AA;
```

```
Query Match 27.0%; Score 1939.5; DB 6; Length 1377;
Best Local Similarity 37.3%; Pred. No. 8.8e-138;
Matches 480; Conservative 202; Mismatches 358; Indels 247; Gaps 39;
Qy 185 IRPENTTVDRIL-ACREEDGELEYLVKVELSYDECYWE-SESIDISTFONEIORF-----237
Db 10 IRPEWMMIIRILNHGFDKKGDVHYLIKDKLPDQCTWEIDIDIPYDNLKQAYWGHRE 69
Qy 238 -----KDVNSRTRSKVDVHKRNPDPFQFDHTPEFLK-----GLLHPYQL 278
Db 70 LMGEDTRLPKLLKKKKLRD-DKQEKPPDTPIDPTVKFDKQWPVIDSTGTLHPYQL 128
Qy 279 EGLNLFRLFSKQTHVILADENGLGHTIQSIALLASLFE--NLIPHVIAPLSTURNWE 336
Db 129 EGLNLFRLFSWAQGTDTILADENGLGHTVQTVIVFLSYLKEGHSKGPYLSAPLSTIINWE 188
Qy 337 REFATWAPQNMVVMVFGTAQARAVIREHEFYLSKQDKIKKKSGGQISSKQKRIKFDV 396
Db 189 REFEMWAPFYVVTYTGDKESRSVIRENEF--SFEDNARISGK--KVFRMKKEVQIKFHV 244
Qy 397 LLTSYEMINLDSAVLKPIKWEQWIVDEGHRLLKNDKSLFSSLTQYSSNHRILLTGTPLQN 456
Db 245 LLTSYELITIDQAILGSIEMACLVVDEARLKNQSKFRVLNSYKIDYKLLLTGTPLQN 304
```


QY 171 -----QMSFNNS-----EDDF--VAIRPEWTVVDRIL--ACREDEGELEYLYVKYKELSY 216
Db 76 DECKSEKRNKKOPLYANKMEERFYRYGKPEWMMHRLNHSFKKGDVHLYKWKDLPY 135
QY 217 DECVCWE--SESDISTFQNEIQRF-----KOVNSRTRRSKQVDHNRNPRDFQ 260
Db 136 DQCTWEIDDDIDIPYDNLKQAYWGHRELMLGEDTRLPKRLKKGKLRD--DKREKPPDTP 194
QY 261 QFHTTPEFLK-----GLLHPYQLEGLNFLRFSWSKOTHVILADEMGLGKTIQSIAL 311
Db 195 IVDPTVKFDQPWYIDSTGTLHPYQLEGLNWLRFSAQGTDTILADEMGLGKTVQTI 254
QY 312 LASIFEE--NLIPHVITAPLTLNWEREFATWAPQMVVYFGTAQARAVIREHEFYLS 369
Db 255 LYSLYKEGHSKGYLVSAPLSTIINWEREFEMAPDFYVVTYTGDKRESRVIRENEFSFG 314
QY 370 KDQKIKKKSGQISSESOKRIKFDVLTSYEMINLDSAVLKPIKWECMIVDEGHLKN 429
Db 315 DNARSKK---VFRMKKEVQIKFHVLLTSYELITIDQAILGSIWACLIVDEAHLKN 370
QY 430 KSKLFSSLTQYSSNHRILITGTLQNNLDELPMFLHFDAGKSGLEEFQEFKINOE 489
Db 371 NQSKFRVLNSYKIDYKLLLTGTPLQNNLELFLNFLTPTFRFNNLEGLFEEFADISKE 430
QY 490 EOLSRHLKMLAPHLRRVKQVWKMPKPELILRVLDLSLOKEYYKAIPTRYQVLYTKK 549
Db 431 DQIKRLHDLGPHMLRLKADVPKFNPAKTELVIRVELSQMKYKFKILTRNFEALNSK 490
QY 550 GGA-QISLNMIMELRVCCHPYV--LEGVE-PVIHDANEAFKOLLESCKQLLDKMMV 605
Db 491 GGGNQVSLNIMDLKCCNHPVLPVAAVEAPVLPNGSNDGSSLYKSSKMLLQKMLK 550
QY 606 KLKEQHRVLIYTOFQHMDDLLEDYCTHKQWYERIDGKVGGAERQIRIDRFNAKSNKF 665
Db 551 KLREDEGRHVLIFSQMTKMLLEDLEFLEYGKYERIDGGITGCLRQEAIDRFNAPGAQOF 610
QY 666 CPLLSTRAGLGINLATADTVIYDSWNPADLOMARAHRLGQTNKWIYLIINRGTI 725
Db 611 CPLLSTRAGLGINLATADTVIYDSWNPADLOMARAHRLGQTNKWIYLIINRGTI 725
QY 726 EERMOLTKKMWLEHLVWCK---LKTQINQBELDDIIRYSGKELFASB----- 772
Db 671 EERITQVAKRKMMLTHLVVRPGLSGSGSKQELDDILKFGTEELFKDDVEGMSQQR 730
QY 773 -----DDEAGSKG-----IHYDDAAIDKLLDRDLVEAE 801
Db 731 PVTPIDVQSSKGNLAASAKKXGSGTTPGDNKQVEDSSVIHYDDAAISKLLDN----- 785
QY 802 EVSVDRDEE---ENGFLKAFKANFVYIDENEAALAEQVAAESKSGNSDRASWEL 858
Db 786 QDATDTELQNMNEYLSFKAQYVVRBEDGVGVEREIIKQBE-----NVD--PDYWEKL 839
QY 859 LKDKFELHQAELNALGKRKRKQL---VSTEEDDLAGLEDVSDGDESY-----EAS 910
Db 840 LRHHYQOQEDLARNLKGKRIKQVNYNDASQDE---WQELSDNQSEYISGEDED 896
QY 911 TDGEAAGQVQTCRRPYRK---GRDNLEPTPLMEGEGRSFVLFQNSQRAIFVQTLAR 967
Db 897 EDPEERPEG-QSQRQSRRLKSDRDKPLP-PLLARVGGNIEVLGFNARQKAFNAIR 954
QY 968 YG-----AGNFDW 975
Db 955 WGMPPQDAFNHW 967
RESULT 13
ID ABB7171 standard; protein; 883 AA.
XX AC ABB7171;
XX DT 26-MAR-2002 (first entry)
XX TT

DE Drosophila melanogaster polypeptide SEQ ID NO 42105.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
KW Drosophila melanogaster.
OS WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR N-PSDB; ABL15874.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Disclosure; SEQ ID NO 42105; 2ipp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB5773-
CC ABB72073). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 883 AA;
SQ

Query Match 24.2%; Score 1738.5; DB 4; Length 883;
Best Local Similarity 42.5%; Pred. No. 8.8e-123;
Matches 377; Conservative 152; Mismatches 256; Indels 103; Gaps 23;
QY 44 RTDAK-----ENACQACGESTNLVSCNCTVAFHAKCLVPLPKDASVENWRCECVSPL 97
Db 24 KTNAKQKFRDEYCKVSDGDLCCDSCPSVYHRTCLSPPLKSTPKGDWICFRCIPLP 83
QY 98 NBIKILDCMRPTKSSSQSSDAEPKPIFVKQYLVKWLKGLSYLHCSWVPEKFEQKAYS 157
Db 84 GRAEKILSRWALDRSVELRTSKGEKR---REYFIKWHGMSYWHCEWIPEGQMLLHAS 139
QY 158 NHRILKTRVNFIRQ-----MESFNNSDDF-----VAIRPEWTVVDRILACRE--DG 203
Db 140 -----MVASFQRRSDMEPSELEDDQDGNLHERFYRYGKPEWLLVQVRINHSEPN 193
QY 204 ELEYLVKYKELSYDECVWESED-ISTFQNEIQRFQKDVNS-----RTRRSKQVDHNRN 256
Db 194 GTWYLVKRELSDYSSWRESDSIPCLNQATALYKLLRSSNKGRRDRPAPTIDLNK-- 251
QY 257 RDPQQPDHTPEFLK--GL-LHPYQLEGLNFLRFSWSKOTHVILADEMGLGKTIQSIAL 313
Db 252 ----KYEDQPVFLKEAGLKLHPFQIEGVSMLRYSWGQGITILADEMGLGKTIQTVVFLY 307
QY 314 SLFPEENLI--PHLVITAPLTLNWEREFATWAPQMVVYFGTAQARAVIREHEFYLSKD 371
Db 308 SLFKEGHCRGPELISVPLSTLTNWERELWAEPLYCVTVYVGKTAARAVIRKEISFEV 367
QY 372 QKKIKKKSGQISSESOKRIKFDVLTSYEMINLDSAVLKPIKWECMIVDEGHLKNKD 431
Db 368 TTKTWRE-----NQYKFNWMLTSYEFISVDAFLGCLDWAALVVDEAHRUNSQ 418

QY 432 SKLPSLSSTOYSSNHRILLTGTPLQNNLDELFLMLHFLDAGKSGLEBFOEFKDINOEEQ 491
DB 419 SKFPRILSKRIGYKLLTGTPLQNNLEELPHLNLFLSSGKFNDLQTFQAEFTDVSKEEQ 478
QY 492 ISRLHKLAPHLARRVKQVDMKPPKELILRLVDLSLQKEYYKAIIFTRNYQVLTKKGG 551
DB 479 VKRLHEILEPHMLRLKADVLKSMPPKSEFIVRVVELSSMQKFYKHLTLTKFKALNQKG 538
QY 552 AQI-SLNNIMELRKVCCHPYMLEGVPEVTHDANEA-----FKQLLESCGKLQLLD 601
DB 539 GRVCSLLNIMDLKCCNHPYLPESA-----ABEATISPSGLYEMSSLTAKSGKLDLS 592
QY 602 KMMVKKEQGHRLVLYTFQFQHMLDLLEDYCTHKKQWYERIDGKVGGAERQIRIDRFNAKN 661
DB 593 KMLKQLKADNRHVLFSQMTKMLNLEHFLGEGYQYDRIDGSLKGLDKQKADRFNDPV 652
QY 662 SNKFCFLSTRAGGLGINLATADTVIYDSDNPHADLQAMARHLRGQTNKVMYRLIN 721
DB 653 SEHFVFLSTRAGGLGINLATADTVIIFSDNPHNDVQAFSRAHRMGQKKVMYRFVT 712
QY 722 RGTIEERMQLTKKMVLHLV-----CKLKTQINQOEELDDIIRYSGKELFASDEDEAG 777
DB 713 HNSVEERIMQVAKHMMMLTHLVVRPGMGW-TTNFSKDELEDILRFGTEDLF-----KDG 766
QY 778 KSGKIHYDDAAIDKLLDRDLVEAEVSVDDDEENGFLKAFKVFANFEYIDENEAALAEAR 837
DB 767 KSEAIHYDDKAVADLLDRTRNGIEE-----KESWANLEYLSSFKVASY-----ATKEDHE 815
QY 838 VAAESSKSAGNSDRASYWEELL-----KDKFELHQAELNALGKR 877
DB 816 EHDYNNDAENTD-PFYWENLMGTVAAPGGRREHGHKGQANPKGNR 862
RESULT 14
ID AAE36105 standard; protein; 1739 AA.
XX AC AAE36105;
XX DT 26-JUN-2003 (first entry)
XX DE Human chromodomain helicase DNA binding protein (CHD) #3.
XX KW Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;
KW chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;
KW cell proliferative disorder; chromatin organisation modifier domain;
KW cytosstatic.
XX OS Homo sapiens.
XX PN W0200298899-A2.
XX PD 12-DEC-2002.
XX PF 03-JUN-2002; 2002WO-US017466.
XX PR 05-JUN-2001; 2001US-0296076P.
XX PR 10-OCT-2001; 2001US-0328605P.
XX PR 22-OCT-2001; 2001US-0338733P.
XX PR 15-FEB-2002; 2002US-0357253P.
XX PR 15-FEB-2002; 2002US-0357600P.
XX PA (EXEL-) EXELIXIS INC.
XX PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
PI Lioubin MN;
XX WPI; 2003-156840/15.
XX PT Identifying a candidate p53 pathway-modulating agent as therapeutic
PT targets for disorders related to defective p53 function e.g. cancer by
PT contacting an assay system having purified CHD polypeptide or nucleic

PT acid, with a test agent.
XX Claim 13; Page 185-192; 278pp; English.
XX CC The present invention relates to a method for identifying candidate p53
CC pathway modulating agents. The method involves contacting an assay system
CC comprising purified chromatin organisation modifier (chromo) domain
CC helicase DNA binding proteins (CHD), nucleic acids, their functionally
CC active fragments or derivatives, with a test agent under conditions
CC where, but for the presence of the test agent, the system provides a
CC reference activity. The methods are useful for identifying modulators of
CC the p53 pathway as therapeutic targets for disorders associated with
CC defective p53 function, such as angiogenic disorders, apoptotic disorders
CC or cell proliferative disorders, e.g. cancer. The modulators are useful
CC as research reagents, diagnostics and therapeutics. The invention is also
CC useful in gene therapy. The present sequence is human chromodomain
CC helicase DNA binding protein
XX
SQ Sequence 1739 AA;
Query Match 22.6%; Score 1621.5; DB 6; Length 1739;
Best Local Similarity 30.7%; Pred. No. 2e-113;
Matches 453; Conservative 229; Mismatches 493; Indels 299; Gaps 43;
QY 11 RSDRKPFVYNLDDSDDDDFVPKK-----DRTFEQVEAIVRTDAKENACQACG 56
DB 199 RGKRRKQDSDDEDDDEAPKQTRRRRAAKNVSYKEDDDE-----TDS-DDLIEMTG 250
QY 57 ESTNLVSCNTCTVAFHAKCLVPLPKDASVENWRCPCVSPNLNIDKILDCMPTKSSEQ 116
DB 251 EGVDEQDQNSSET-----IEKVLDD--SRLGKKGAT 277
QY 117 GSS-----DAEKPPIFVK-----OVLVWKGLSYLHCSWVPEKFEQ-----KAYS 157
DB 278 GASTTVYAEANGDPSGDFDTEKDEGEIQYLKWKGSWISHSWEESESLOQKVKGLKX 337
QY 158 NHRKLRVNNFHR-----QMESFNNGEDDFVAIRPEWTTVDRILACREE-----TRR 246
DB 338 LENFKKKEDEIKQWLGKVPEDVEYFCQELASELNKQVIVVERVIAVTKSTKLQTD 397
QY 202 -----DGELEVLYVKYKELSYDECYWESESDI-STFONEIQRFKDVNSR-----TRR 246
DB 398 FPAHSRKPAPSNEPEYLCKWMGLPYSECSWEDSALIGKPKQNCIDSFHSRNSKTIPTRE 457
QY 247 SKDVIDHKNRPDRDQFQDHPPEFLKG--LHPYQLEGLNPLRPSWSKQTHVILADEMGLG 303
DB 458 CKAL--KQRP--FVALKKQPAYLGGENLELDYQLEGLNLWLAHCKNSVILADEMGLG 514
QY 304 KTIQSIALLASLFEENLI--PHLVIAPLSLRNWEREFATWAPQMVVYFGTAQARAVI 361
DB 515 KTIQTISFLSYLPHQHLQVGFPLIVVPLSTLTSWQREFEWAPEINNVVYIGDLSRNTI 574
QY 362 REHEFYLSKDKKIKKKKSGQISSBSKQRIKFDVLTLSYEMINLDSAVLKPIKWECMIV 421
DB 575 REYEWI-----HSQTKLKFENALITTYEILLKDKTVLGSINWAEFLG 616
QY 422 DEGHRLKNKDSKLFSSILTOYSSNHRILLTGTPLQNNLDELFLMLHFLDAGKSGLEBFOE 481
DB 617 DEAHRLNDDSLYKTLIDPFKSNHRLITGTPLQNSLKELWSLHLFIMPEKFEWEDFEE 676
QY 482 EFKDINQEEQISRLHKMLAPHLRLRRVKQVDMKPPKELILRLVDLSLQKEYYKAIIFTR 541
DB 677 DHGK-GRENGYQSLHKVLEPFLRLRRVKQVDEKSLPAKVEQILRVENSALQKYVKKWILTR 735
QY 542 NYQVLT-KGGAQISLNNIMELRKVCCHPYMLEGVPE-VIHDANEAFAKLLSCKGLQL 599
DB 736 NYKALAKTRGSTSGFLNIVMELKCCNHCYLLKPPPEENERENGQETLLSLIRSSGKLL 795
QY 600 LDKMMVKLEQGHRLVLYTFQFQHMLDLLEDYCTHKKQWYERIDGKVGGAERQIRIDRFNA 659
DB 796 LDKLLTRLRERGNRVLIIFSQWVEMLDILAEYLIKHYPFORLDGSIKGEIRKQALDHFA 855
QY 660 KNSNKFCLLSTRAGGLGINLATADTVIYDSDNPHADLQAMARHLRGQTNKVMYRL 719

```

Db      856  DGEDFCFLSTRAGGLINLASADTVIFDSWNFONDQAQRAHRRGQKKQVNIYRL 915
Qy      720  INRGTIIEERMOLTKKMWLEHLV-----GKLTQNIINOEELEDDIIRYX 764
Db      916  VTGKTVEEETIERAKKMWLDHLVIQKMDITGRTILENNSGRSGNSPNFKEELTAILKFG 975
Qy      765  SKELFASDEDEAGSKGIHYDDAIDKLLDRDLVVEAEVSVDEEENGFLKAFKVANFEY 824
Db      976  AEDLFKELEGESEP-----QEMDIDEIL--RLAETRENEVTSATDELSQPKVANPAT 1028
Qy      825  IDENEAALAEAOORVAASKSSAGNSDRASYEELL-----KDKFELHQAELNALGK 876
Db      1029  MEDEE--ELEER-----PHKD-----WDIIPEEQKVKVEERQKELEEIYMLPR 1072
Qy      877  RKRSRQLVSIIEBDDLAGLEDVSDGDESYAE-STDGEAAGQGVQTRRPRYRKGRDNL 935
Db      1073  IRSSTKK-----AQTNDSDSTESKQQAQSSASESETSDDDKKPKER----- 1117
Qy      936  EPTPLMEGGRSFR-----VLGFNOSQRAIFVOTLMRYGAGNFDWKEFVPR---LKQKTPEE 989
Db      1118  -----GRPSVRKOLVEGFTDAEIRRFKAYKFGI--PLERLECLARDAELVDSKSVAD 1169
Qy      990  INEYGILF-----LKHIAEEDIDENSPTFSDGVPKEGLRIEDVLVRIALLILVQEKVXF 1042
Db      1170  LKRLGELIHNCSVAMQVEEEOQENASEGKPGKRRGPTIKISGVQVNVKSIIOHEEBF 1229
Qy      1043  VEDHPGKPPPSR-----ILERPFLGRSGKIWEKHEHDKIMIRAVLKHGYGRWQAIVDDKE 1097
Db      1230  EMLHKSIPVDPBEKKYCLTCRYKAAHFDVWGVGVEDDSRLLLGIYEHGYGNELIKTDPE 1289
Qy      1098  LGTQELIC-----KELNPFHISLSAAEQAGLQGGQSGSGSNPGAGTNPNGSVIT 1147
Db      1290  LKLTDLKILPVETDKPGQKQLOTRADYLLKLLKGLKK-----GAVT 1332
Qy      1148  GNNASADGQVNSMFYRDMORRILFEVKKRVLLLEKAMWYEAEEYVGLGSSSIPT 1207
Db      1333  GGEAK-----LKKRVRVKKENKVPRLKEEHGIELSSPRHSD 1370
Qy      1208  EPEAEPIADTVGVSPFIEVDDEMDGLPKTDPTITSBEIMGAAVDNNQARVETAOHYNQMC 1267
Db      1371  NPSEGEVKD-----DGLEKS--PMKKQKKKENKEN-----KQMS 1406
Qy      1268  KLLD-ENARESQAQVNNPPSTKYNESFRALKSINGNINTILSITSDOSKSHEDTKPD 1326
Db      1407  SRDKEGDKERKSKDKKPKSGDAKSSKRSQGPVHITAGSPVPTGEDEDD---D 1463
Qy      1327  LNNVEMKDTAEETKPLRGVVDLNVVEGENIAE 1360
Db      1464  LDQETFSICKERMVRPKKALKQDKPKGLNVQE 1497

```

RESULT 15

```

AAE36106
ID  AAE36106 standard; protein; 1739 AA.
XX
AC  AAE36106;
XX
DT  26-JUN-2003 (first entry)
XX
DE  Human chromodomain helicase DNA binding protein (CHD) #4.
XX
KW  Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;
KW  chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;
KW  cell proliferative disorder; chromatin organisation modifier domain;
KW  cytostatic.
XX
OS  Homo sapiens.
XX
FH  Key Location/Qualifiers
FT  Domain 281..327
FT  /note= "Chromodomain"
FT  375..433
FT  Domain

```

```

FT      11  RSDRKPVNLDSDDDDFVPKK-----DRTFEQVEAIVRTDAKENAQACG 56
FT      199  RGRKKQDSSDEDDDDDEAPKQTRRAAKNVSYKEDDDFE-----TDS-DDLEMTG 250
Qy      57  ESTNLVSCNTCTYAFHAKCLVPLPKDASVENWRCPECVSPPLNEIDKILDCENRPTKSSQ 116
Db      251  EGVDEQQDNSET-----IEKVLQ--SRLGKKGAT 277
Qy      117  GSS-----DAEPKPIPVK-----QYLKVKWGLSYLHCSWVPEKFEQ-----KAYKS 157
Db      278  GASTTYAIEANGDPSPGDFTEKDEGEIQYLKWKWGSYIHSTWSEESLQOQKVGLKK 337
Qy      158  NHRLLKTRVNNFHR-----QWESFNNSDDFVAIRPEWTTVDRLACREE----- 201
Db      338  LENFKKDEDEIKQWLKGVSPEDVEYFNCCQQLASELNKQIQUVERVIAVTKTSKSTLGQTD 397
Qy      202  -----DGELEYLVKYSYDECYWSESDI-STFQNEIQRFKDVNSR-----TRR 246
Db      398  FPAHSRKPAPSNEPEYLCKWMLPYSECSWEDEALIGKFKQCIDSFHRSNNKTIPTRE 457
Qy      247  SKDVHKNRPDFOQFDHTPEFLKG---LLHPYQLEGLNLFRLFSWSKQTHVILADEMGLG 303

```

Query Match 22.6%; Score 1621.5; DB 6; Length 1739;
 Best Local Similarity 30.7%; Pred. No. 28-113;
 Matches 453; Conservative 229; Mismatches 493; Indels 299; Gaps 43;

```

Qy      11  RSDRKPVNLDSDDDDFVPKK-----DRTFEQVEAIVRTDAKENAQACG 56
Db      199  RGRKKQDSSDEDDDDDEAPKQTRRAAKNVSYKEDDDFE-----TDS-DDLEMTG 250
Qy      57  ESTNLVSCNTCTYAFHAKCLVPLPKDASVENWRCPECVSPPLNEIDKILDCENRPTKSSQ 116
Db      251  EGVDEQQDNSET-----IEKVLQ--SRLGKKGAT 277
Qy      117  GSS-----DAEPKPIPVK-----QYLKVKWGLSYLHCSWVPEKFEQ-----KAYKS 157
Db      278  GASTTYAIEANGDPSPGDFTEKDEGEIQYLKWKWGSYIHSTWSEESLQOQKVGLKK 337
Qy      158  NHRLLKTRVNNFHR-----QWESFNNSDDFVAIRPEWTTVDRLACREE----- 201
Db      338  LENFKKDEDEIKQWLKGVSPEDVEYFNCCQQLASELNKQIQUVERVIAVTKTSKSTLGQTD 397
Qy      202  -----DGELEYLVKYSYDECYWSESDI-STFQNEIQRFKDVNSR-----TRR 246
Db      398  FPAHSRKPAPSNEPEYLCKWMLPYSECSWEDEALIGKFKQCIDSFHRSNNKTIPTRE 457
Qy      247  SKDVHKNRPDFOQFDHTPEFLKG---LLHPYQLEGLNLFRLFSWSKQTHVILADEMGLG 303

```

/note= "Chromodomain"
 487..768
 /note= "SNF2N domain"
 831..905
 /note= "Conserved C-terminal domain"
 WO200298899-A2.
 12-DEC-2002.
 03-JUN-2002; 2002WO-US017466.
 05-JUN-2001; 2001US-0296076P.
 10-OCT-2001; 2001US-0328605P.
 22-OCT-2001; 2001US-0338733P.
 15-FEB-2002; 2002US-0357253P.
 15-FEB-2002; 2002US-0357600P.
 (EXEL-) EXELIXIS INC.
 Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
 Lioubin MN;
 WPI; 2003-156840/15.
 Identifying a candidate p53 pathway-modulating agent as therapeutic
 targets for disorders related to defective p53 function e.g. cancer by
 contacting an assay system having purified CHD polypeptide or nucleic
 acid, with a test agent.
 Claim 13; Page 192-199; 278pp; English.

The present invention relates to a method for identifying candidate p53
 pathway modulating agents. The method involves contacting an assay system
 comprising purified chromatin organisation modifier (chromo) domain
 helicase DNA binding proteins (CHD), nucleic acids, their functionally
 active fragments or derivatives, with a test agent under conditions
 where, but for the presence of the test agent, the system provides a
 reference activity. The methods are useful for identifying modulators of
 the p53 pathway as therapeutic targets for disorders associated with
 defective p53 function, such as angiogenic disorders, apoptotic disorders
 or cell proliferative disorders, e.g. cancer. The modulators are useful
 as research reagents, diagnostics and therapeutics. The invention is also
 useful in gene therapy. The present sequence is human chromodomain
 helicase DNA binding protein

Sequence 1739 AA;

Db 458 CKAL--KORPR-FVALKKPAYLGGENLELDYQLEGLANLWLAHWCNNKNSVILADEMGLG 514
Qy 304 KTQSIALLASLEENLI--PHLVIAPLSTLRWEREFATWAPQNVVYFCTQAARAVI 361
Db 515 KTIQTISFLSFHQHQLYGFPLIWPPLSTLSMQREFIWAPEINVVYIGDLMRSNTI 574
Qy 362 REHEFYLSKQKTIKKKSGQISSESKORIKFVLLTSYEMINLDSAVILKPIKWECMIV 421
Db 575 REYEWI-----HSQTKLKNALITYEILLKDKTVLGSINWAFGLV 616
Qy 422 DEHRLKNKDSKLFSSITQYSSNHRILLTGTPLONNLDELFLMLHFLDAGKFGSLEEFQE 481
Db 617 DEARLKNDDSLLYKTLIDFKSNHRLITGTPLQNSLKLWLSLHFTIMPEKFEFWEDFEE 676
Qy 482 EFKDINQEEQISRLHKMLAHLRLRVKQVMDKMPKPKELILRVDLSSLOKEYYKAIFTR 541
Db 677 DHGK-GRENGYQSLHKLEPFLRLRRVKDVEKSLPAKVEQILRVEMSAHQYKWLTR 735
Qy 542 NYQVLT-KGGAQISLNNIMELRKVCCHPVMLEGVPR-VIHDAEAFKOLLESCEGKLQ 599
Db 736 NYKALAGTRGTSGLPNIWELKCCNHCYLIKPEENERENGQEILLSLIRSSGKLIL 795
Qy 600 LDKMMVKLQCGHRLIYTOQHMLDLLEDYCTHKQWYERIDGKVGGAERQIRIDRFNA 659
Db 796 LDKLLTLRERGNRVLIFSQWVRMLDILAEVLTIKHYPPQRLDGSIKGEIRKQALDHFA 855
Qy 660 KNSKFCFLSTRAGGGINLATADTVIYVDSDNPHADLOAMARHRLGQTNKVMYRL 719
Db 856 DGSDFCFLSTRAGGGINLASADTVIYFSDWNPNQDLQAQARAHRIQKKQVNIYRL 915
Qy 720 INRGTEERMMOLTKKQWLEHLV-----GKLTQINQBELDDIIRYG 764
Db 916 VTKGTVEEIIERAKKQWLDHLVIQRMDDTGTTRILENNSGRSNSNPFNKEELTAILKFG 975
Qy 765 SKELFASDEDPAGSKGIHYDDAIDKLDRDLVEABEVSDDEENGFLKAFKANFEY 824
Db 976 AEDLFKELEGESEP-----QEMDIDEL--RLAETRENEVSTSATDELLSQPKVANFAT 1028
Qy 825 IDENEAALAEQORVAERKSAGNSDRASWEELL-----KDFELHQAELNALGK 876
Db 1029 MEDEE--ELEER-----PHKD-----WDEIIPBEQRKVEEERQKELEBIYMLPR 1072
Qy 877 RKRSRKOLVSTEEDDLAGLEVDSDGDESVEAE--STDGEAAGQVQTRRPRYRKGRDNL 935
Db 1073 IRSSTK-----AQTNDSDSDESKQAQSSASESETEDSDDDKKPKR----- 1117
Qy 936 EPTPLMEGEGRSFR---VLGNQSORAIFVOTLMRYGAGNFDWKEFVPR---LKQKTFFEE 989
Db 1118 -----GRPRSVRKDLVEGFTDAEIRRFKAYKKFGL-PLERLECLARDAELVDKSVAD 1169
Qy 990 INEYGILF-----LKHIAEIDENSPTSDGVPKESGLRIEDVLVRIALLILVOEKVKF 1042
Db 1170 LKRLGELTHNSCVAMQEEYEQLENASEGKPGKRRGPTIKITSGOVNVKSIIOHBEFF 1229
Qy 1043 VEDHPGKPVPPSR-----ILERFFGLRSKGIWKEEHDKIMIRAVLKHGGRQWQAIVDDKE 1097
Db 1230 EMLHKSIPVDEEKKYCLTCRVKAAHFVWGVVEDDSRLLLGIVHEGYGNWELIKTDPE 1289
Qy 1098 LGIQELIC-----KELNFPHI SLSAAEQAGLOGQNGSGSNPGAQTNQNGPSVIT 1147
Db 1290 LKLTDKILPVETDKKPOGQLQTRADYLLKLLRKGLEKK-----GAVT 1332
Qy 1148 GNNNASDAQVNSMFYFRDMORLIVFVKRVLLLEKAMNYEABEYVGLGSSSIPTTE 1207
Db 1333 GGEBAK-----LKKRKPVRVKENKVPRLKEEHGIELSSPRHSD 1370
Qy 1208 EPEAEPKIADTVGVSFIEVDDMDLGLPKTDPTITSEEIMGAADVNOARVEIAQHYNQMC 1267
Db 1371 NPSEGEVKD-----DGLEKS--PMKKKQKKKENKE-----KQMS 1406
Qy 1268 KLLD-ENARESVQAYVNNQPPSTKVNESFRALKINGNINTILSITSDQSKSHEDDTKPD 1326

Db 1407 SRKDKSGDKERKSKDKKPKSGDAKSSKSKRSQGPVHITAGSEVPVIGEDEDD---D 1463
Qy 1327 LNNVEMKDTAEETKPLRGVGVVDLNVVEGEBENIAE 1360
Db 1464 LDQETFSICKERNRPVKKALKQLDKPKGLNVQE 1497

Search completed: September 14, 2004, 02:04:22
Job time : 173 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 14, 2004, 02:01:31 ; Search time 36 Seconds
(without alignments)
1984.732 Million cell updates/sec

Title: US-10-049-137-2
Perfect score: 7187
Sequence: 1 MSLSVRLRLRSRDKPYNL.....VDVMEEAKEBEKPMNVVD 1384

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/prodata/2/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2091	29.1	1912	4	US-08-913-832A-2
2	2091	29.1	1912	4	US-09-249-181A-2
3	2091	29.1	1912	4	US-09-158-707-2
4	1623.5	22.6	1739	4	US-09-976-594-76
5	1028.5	14.3	1646	4	US-09-535-008-67
6	1027	14.3	1647	4	US-09-535-008-2
7	1027	14.3	1649	4	US-09-535-008-75
8	1025.5	14.3	1650	4	US-09-535-008-71
9	1011	14.1	1681	4	US-09-535-008-77
10	1009.5	14.0	1678	4	US-09-535-008-69
11	1009.5	14.0	1682	4	US-09-535-008-73
12	1008	14.0	1679	4	US-09-535-008-65
13	997	13.9	831	4	US-09-702-705-1819
14	997	13.9	831	4	US-09-736-457-1819
15	997	13.9	831	4	US-09-671-325-1819
16	898.5	12.5	801	3	US-09-104-070-2
17	699.5	9.7	904	4	US-09-252-991A-29119
18	648	9.0	1070	4	US-09-107-532A-3716
19	620	8.6	1073	4	US-09-134-000C-6356
20	619.5	8.6	1247	4	US-09-404-627-2
21	537	7.5	2972	4	US-09-579-181-2
22	537	7.5	3118	4	US-09-579-181-1
23	459	6.4	1075	4	US-09-198-452A-916
24	402.5	5.6	727	3	US-09-179-558-56
25	402.5	5.6	727	4	US-09-722-825-56
26	402.5	5.6	727	4	US-09-722-487-56
27	402.5	5.6	727	4	US-09-722-708-56

28	402.5	5.6	941	3	US-09-179-558-55	Sequence 55, Appl
29	402.5	5.6	941	4	US-09-722-825-55	Sequence 55, Appl
30	402.5	5.6	941	4	US-09-722-487-55	Sequence 55, Appl
31	402.5	5.6	941	4	US-09-722-708-55	Sequence 55, Appl
32	402	5.6	949	4	US-09-404-627-4	Sequence 4, Appl
33	331	4.6	946	4	US-09-252-991A-18989	Sequence 18989, A
34	318	4.4	722	4	US-09-252-991A-26452	Sequence 26452, A
35	306.5	4.3	459	4	US-09-134-000C-6639	Sequence 6639, Ap
36	241	3.4	301	4	US-09-198-452A-898	Sequence 898, App
37	239	3.3	359	3	US-09-179-558-65	Sequence 65, Appl
38	239	3.3	359	4	US-09-722-825-65	Sequence 65, Appl
39	239	3.3	359	4	US-09-722-487-65	Sequence 65, Appl
40	239	3.3	359	4	US-09-722-708-65	Sequence 65, Appl
41	233.5	3.2	951	4	US-09-328-352-4456	Sequence 4456, Ap
42	220.5	3.1	247	4	US-09-370-838-112	Sequence 112, App
43	220.5	3.1	1003	4	US-09-489-039A-12357	Sequence 12357, A
44	220	3.1	351	4	US-07-945-295-2	Sequence 2, Appl
45	220	3.1	351	5	PCT-US91-06418-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-913-832A-2
; Sequence 2, Application US/08913832A
; Patent No. 6329517
; GENERAL INFORMATION:
; APPLICANT: Seelig, Hans Peter
; APPLICANT: Renz, Manfred
; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
; FILE REFERENCE: 8484-0030-999
; CURRENT APPLICATION NUMBER: US/08/913,832A
; CURRENT FILING DATE: 1998-01-12
; PRIOR APPLICATION NUMBER: PCT/DE96/00444
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1912
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-913-832A-2

Query Match	29.1%	Score 2091;	DB 4;	Length 1912;
Best Local Similarity	35.5%;	Pred. No. 8.3e-166;		
Matches 520;	Conservative 205;	Mismatches 424;	Indels 314;	Gaps 44;
QY	21	DSDDDDDFVKDRTFEQVEAIVRTDAKENACCGESTNLVSCNTCTYAFHAKCLVPL 80		
DB	426	DNSEGEIILEEVGGDLEED-----DHHMEFCRVCKDGGELCCDCTCPSSYHCHLNPL 480		
QY	81	KDASVENWCPECVSPL--NEIDKILDCB--NRPTKSSEQSSDAE-----PKPIF---V 128		
DB	481	PEIPNGEWLCPRCTCPALGKQKILIKWQGPSTPTVPRPPDADPTSPPLSGRPE 540		
QY	129	KQYLKWKGLSYLHCSWPKE-----FQKAYKSNHRLKTRVNNFRHQB----- 173		
DB	541	RQFFVKQMGWSYHCSWVSELQLELHCQVMFRYQKNDMDPEPPSGDFGDEKSRKKN 600		
QY	174	---SFNNSDDPF--VAIRPWTVDRIIL-ACREDEGELEYLYKYLKSLVDECYWESF--- 224		
DB	601	KDPKFAEMEEFRYRGIKPEWMMIHLNHSVDKGGHVLTKWRDLDPYDQASWESDEVE 660		
QY	225	-SDISTFONEIQRFKDV-----NSRTRRSKVDVHKENPR-----DFQOFDHTPEFLK 270		
DB	661	IQDYDLFKOSYNNHRLMGEGRPKKKVKLKLRLPPETPTVPTVKYERQREYLD 720		
QY	271	---GLLHPYQLEGLNPLRFSWKSQTHVILADEMGLKTIQSIALLASLFEF---NLIPHLV 325		
DB	721	ATGGTLHPYQMEGLNPLRFSWKSQTHVILADEMGLKTIQSIALLASLFEF---NLIPHLV 325		
QY	326	IAPLSTLRNWEREFATWAFQMNVMVFTGAQARAVIREHEFYLSKQDKIKKKKSGQISS 395		

```
Db 781 SAPLSTIINWEREFEMWAPDMYVTVYVGDKDSRAIIRENEFSPEDNAIRGGKASRM--- 837
Qy 386 ESQKRIKFDVLTSTYEMINLDSAVLKPIKWECHIVDEGHRLLKKNKSKLFSSITQYSSNH 445
Db 838 -KKEASVXFHVLTSYELITIDMAILGSDIWAFLVDEAHLKKNQSKFRVLNGYSLOH 896
Qy 446 RILLTGTPLONNDELPMHFLDAGKFGSLEEFQEFKDINOBEQISRLHKMLAPHLLR 505
Db 897 KLLTGTPLQNNLEELFHLNFLTTPFRHNLGFLFEEFADIAKEDQIKKLHMLGPHMLR 956
Qy 506 RVKDKVMKMPKKEKILRLVDLSLOKEYYKAIPTFRNYQVLTTKGGA-QISLNNIMMELR 564
Db 957 RLKADVFKMPSKTEILVRVELSPMQKYYKIILTRNFALNARGGQVSLNNVMDLK 1016
Qy 565 KVCCHPYMLGVEPVTHDANEAFKQ-----LLESCCKLQLLDKMMVKLEQGHVRL 615
Db 1017 KCCNHPYLF----PVA--AMEAPKPNMGMYDGSALIRASGKLLLLQKMLKKEGGHRVL 1070
Qy 616 IYTOFQHMLDLLEDYCTHKKWOYERIDGKVGGAERQIRIDRENAKNSKFCFLSTRAGG 675
Db 1071 IFSQMTKMLDLEDFLEHGKYEYRIDGGITGNMRQEAIDRFNAPGAQFCFLSTRAGG 1130
Qy 676 LGINLATADTVIYDSDWNPHADLQAMARAHRLGQTNKVMYIRLINRGTTIEERMQLTKK 735
Db 1131 LGINLATADTVIYDSDWNPENDIQAFSRAHRIQGNKVMYIRFVTRASVEERITQVAKK 1190
Qy 736 KMULEHLVWGK--LKTQNTNQBELDIIIRYGSKELFASDEDDAGKSGK-----IHYD 785
Db 1191 KMWLTHLVVRPGLSGTKGSMKSQBLDILKFGTEELFKEDATDGGGDNKEGDSVVIHYD 1250
Qy 786 DAAIDKLLDRLDVEABEVSVDDEENGFLKAFKVFANFEYIDENEAALAEQRAVAESKSS 845
Db 1251 DKALERLLDNRQDTEDEL--QGMNEYLSFPKVAQVVRREEMGEVEEREIIKQES 1308
Qy 846 AGNSDRASYEELKQFELHQAELNALGKRKRKQL-----VSIBEDDLAGLEDVSSD 901
Db 1309 VD---PDYMEKLLRHVEQQEEDLARNLGKGRIRKQVNVNDGSGQEDRD---WQDDQSD 1361
Qy 902 GDESYEABTDG--EAAQCGVQTGRPYR---RKGRDNLEBTPMLMEGRGRFVLGFMQSQ 957
Db 1362 NQSDYSVASEEGDDFERSEAPRPSRKGLRNDKOKPLP--PLLARVCGNIEVLGNFARQ 1420
Qy 958 RAIFVQTLRMVG-----AGNFDNKEFVPRLKOKTFEINEYGIILFLAHIAEEDENSPTF 1012
Db 1421 RKAFENAIMRYGMPQDAFTQW--LVDRLGKSEKFKAYVSLFMRHLCEPGADGAETP 1478
Qy 1013 SDGVPKEGLRIEDVLRIALLILVOEKVKFVEDHPGKVPFPFPRILFRPGLRSGKIWKEE 1072
Db 1479 ADGVPREGLSRQHVLTRIGVMSLIRKKVQBFE-----SMP--ELAEVEENKMSQPSG--P 1536
Qy 1511 -----HVNGRW-----TQNT----- 1510
Qy 1073 HDKIMIRAVLKHGVRQQAIVDDKELGIGBELIKELNFPHPHISLSAABQAGLOGQNGSGGS 1132
Db 1511 -----HVNGRW-----SMP--ELAEVEENKMSQPSG--P 1536
Qy 1133 NPGAQTNQPCSVITGNNASADGAQVNSMFFYRDMQRRLVEFKVRKVLLEKAMNYEYA 1192
Db 1537 SPKPTPTSTPGD-----TQNT----- 1553
Qy 1193 BEYYGLGSSSIPTPEPEAEPIADTVGSVFIEVDDMLDGLPKPTDPTSEEINGAAVDN 1252
Db 1554 -----PAPVPPAE-----DGIK----- 1565
Qy 1253 NOARVEIAQYNQCKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNI--NTILSI 1311
Db 1566 -----IEEN-----SLKEESIERGEKEVSTAPETAIEC 1594
Qy 1312 TSDQSKSHEDDT-----KPDLLNNVEMKDTAB---ETKPLRGVGVDLNVAVEEENIAE 1360
Db 1595 TQAPAPASEKVVVEPEGEKEVEKAEVKERTPEETEP--KGAADVKEV--EKSAID 1651
Qy 1361 ASGSVDVMEBAEKEEKPKNMV 1383
```

```
Db 1652 LTPIVVEDKEEKEEBEKEVML 1674
RESULT 2
US-09-249-181A-2
; Sequence 2, Application US/09249181A
; Patent No. 640679
; GENERAL INFORMATION:
; APPLICANT: Seelig, Hans Peter
; APPLICANT: Renz, Manfred
; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
; FILE REFERENCE: 8484-0059-999
; CURRENT APPLICATION NUMBER: US/09/249,181A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/913,832
; PRIOR FILING DATE: 1998-01-12
; PRIOR APPLICATION NUMBER: PCT/DE96/00444
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1912
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-249-181A-2
Query Match 29.1%; Score 2091; DB 4; Length 1912;
Best Local Similarity 35.5%; Pred. No. 8.3e-166; Indels 314; Gaps 44;
Matches 520; Conservative 205; Mismatches 424;
Qy 21 DSDDDDFVPKDDRTFEQVEAIVRTDAKENACQAGESTNLVSCNTCTYAFHAKCLVPL 80
Db 426 DNSEGEILLEVGDLDEED----DHFMEFCRVCKDGGELLCCDTCPSYHIHCLNPPL 480
Qy 81 KDAENVNWCPECVSP--NEIDKILDCR--MRPTKSSQSGSDAE-----PKDIP--V 128
Db 481 PEIPNGEWLCPRCTCALGKVKQILIKWQGPSPPTVPRPPDADPNTSPKLEGRPE 540
Qy 129 KOYLVKWKGLSYLHCSWVPEK-----FQKAYKSNHRLKTRVNFHQQME----- 173
Db 541 RQFVVKWQGSYWHCSWSVSELQLEHCQVNFYQKNDMDPEPSPGDFGDEKSKRKN 600
Qy 174 ---SPNSEDDF--VAIRPETWTTVDRIIL--ACREDEGELEYLVKYKLSYDECYWESE--- 224
Db 601 KDPKFAEMEERFVRYGKPEWMMIHLNHSVDKKGHVHYLIKWRDLPYDQASWESEDVE 660
Qy 225 -SDISTFQNEIQRFKV-----NSRTRSKVDVHKRNP-----DFQQFHTPEFLK 270
Db 661 IQDYDLFKQSYWNHRELMRGEGRPGKCLKVKLRLEPPETPTVDPTVKYERQPEYLD 720
Qy 271 ---GLLHPYOLEGLNFRFSWSKOTHVILADEMGLGKTIQSIALLASLPEE--NLIPHLV 325
Db 721 ATGGLTHPYQMEGLNMLRFSWAOGTTILADEMGLGKTVQTAFLYSLYKEGHSKGFFLV 780
Qy 326 IAPLTLRNWEREFATWAPQNMVYVFGTAQARAVIREHEFYLSKDQKKIKKKKSGQISS 385
Db 781 SAPLSTIINWEREFEMWAPDMYVTVYVGDKDSRAIIRENEFSPEDNAIRGGKASRM--- 837
Qy 386 ESQKRIKFDVLTSTYEMINLDSAVLKPIKWECHIVDEGHRLLKKNKSKLFSSITQYSSNH 445
Db 838 -KKEASVXFHVLTSYELITIDMAILGSDIWAFLVDEAHLKKNQSKFRVLNGYSLOH 896
Qy 446 RILLTGTPLONNDELPMHFLDAGKFGSLEEFQEFKDINOBEQISRLHKMLAPHLLR 505
Db 897 KLLTGTPLQNNLEELFHLNFLTTPFRHNLGFLFEEFADIAKEDQIKKLHMLGPHMLR 956
Qy 506 RVKDKVMKMPKKEKILRLVDLSLOKEYYKAIPTFRNYQVLTTKGGA-QISLNNIMMELR 564
Db 957 RLKADVFKMPSKTEILVRVELSPMQKYYKIILTRNFALNARGGQVSLNNVMDLK 1016
Qy 565 KVCCHPYMLGVEPVTHDANEAFKQ-----LLESCCKLQLLDKMMVKLEQGHVRL 615
Db 1017 KCCNHPYLF----PVA--AMEAPKPNMGMYDGSALIRASGKLLLLQKMLKKEGGHRVL 1070
```

```
QY 616 IYTOFQHMULDLEDYCTHKKWOYERIDGKVGGAERQIRIDRFNAKSNKFCFLLSTRAGG 675
Db 1071 IFSQMTKMLDLEDFLEHEGYKERYIDGGITGNMQEADIRFNAPGAQQFCFLLSTRAGG 1130
QY 676 LGINLATADTVIYDSDWNPADLQAMARAHRLGQTNKVMYIRLNRGTIEERMMLTKK 735
Db 1131 LGINLATADTVIYDSDWNPADLQAFSAHRIGQNKVMYIRFVTRASVEERITQVAKK 1190
QY 736 KMVLEHLVVGK---LKTQINQBELDDIIRYGSKELFASEDDBAGSKG-----IHYD 785
Db 1191 KMMLTHLVVRPGLSGTSGMSKQELDDILKFGTEELFKDEATDGGDNKEGEDSSVIHYD 1250
QY 786 DAAIDKLLDRDLVBAEVSVDDEENGFLKAFKVFANFEYIDENEAALAEQORVAASKSS 845
Db 1251 DKAIERLLDRNQDETDTEL--QGMNEYLSFSFVAQYVVRSEEMGEERIEIKQES 1308
QY 846 AGNSDRASWEELKDKFELHQAELNALGKRKRKQL-----VSTEEDDLGLAEVDSSD 901
Db 1309 VD-----PDYWEKLLRHHEYOQOEDLARNLGKGRIRKQVNYNDGSDQDRD---WQDDQSD 1361
QY 902 GDSYEAESTDG-EAAGQGGVQTGRPYR---RKGRDNLPTPLMEGEGRSFRVLGNQSQ 957
Db 1362 NQSDYSVASEGEDDFERSEAPRPSRKLGRNDKPLP-PLLARVGGNIEVLGNARQ 1420
QY 958 RAIFVQTLARYG-----AGNFDWKEFVPRLKQKTFEINEYGLFLKHAIEIDENSPTF 1012
Db 1421 RKAFLNAINRYGMPQDAFTQW--LVRLDGRGSEKEFRAYVSLFMRHLCEPGADGAET 1478
QY 1013 SDGVPKEGLRIEDVLAIRIALLIIVQEKVFEVDHPGKVPVPPSILRERFFCLRSKGKWKSE 1072
Db 1479 ADGVPREGLSRQHVLTRIGVMSLIRKKVOE-----1510
QY 1073 HDKIMIRAVLKHGYSRQWAIYDVKELGLOBLICKELNFPHISLSAAEQAGLOQSGSGS 1132
Db 1511 -----HNGRW-----SMP--ELAEVEENKMSQPS--P 1536
QY 1133 NPGAQTNQNGSVITGNNSADGAQVNSFYRDMQRRLVEFVKRVLLEKAMNYEVA 1192
Db 1537 SPKTPFTPGD-----TOPNT-----1553
QY 1193 EBYGLGSSSIPTERPEAPKIDTGVGSFIEVDDEMDLGLPKDPTITSEIMGAVDN 1252
Db 1554 -----PAPVPAE-----DGIK-----1565
QY 1253 NQARVEIAQHYNQCKLLDENARESQAQVYVNNOPSTKVNESFRALKSINGNI-NTILSI 1311
Db 1566 -----IEEN-----SIKEEESIEGEKEVKSTAPETAIEC 1594
QY 1312 TSDQSKSHEDDT-----KPDNLNVEMKDTAE---ETKPLRGVVDVNLNVVEGENIAE 1360
Db 1595 TQAPAPASEDEKVVVEPPEGEKEVEKAEVKERTEEPMETEP--KGAADVEKVB-EKSAID 1651
QY 1361 ASGSVDVWEEAKEEEKPKMVV 1383
Db 1652 LTPIVVEDKEEKEEKEEKEVWL 1674
```

RESULT 3

```
US-09-158-707-2
; Sequence 2, Application US/09158707
; Patent No. 6500923
; GENERAL INFORMATION:
; APPLICANT: Seelig, Hans Peter
; APPLICANT: Renz, Manfred
; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
; FILE REFERENCES: 8484-0043-999
; CURRENT APPLICATION NUMBER: US/09/158, 707
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0.
; SEQ ID NO 2
; LENGTH: 1912
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-158-707-2
```

```
Query Match 29.1%; Score 2091; DB 4; Length 1912;
Best Local Similarity 35.5%; Pred. No. 8.3e-166;
Matches 520; Conservative 205; Mismatches 424; Indels 314; Gaps 44;
QY 21 DSDDDDDVFPVKDRITFEQVEAIVRTDAKENACQAGESTNLVSCNTCTTAFHAKCLVPLP 80
Db 426 DNSEGEIILEEVGGDLEED-----DHHMEFCRVCKDGGGELLCCDCTCPSSYIHCLNPPL 480
QY 81 KDAIVENRCPCVPL--NEIDKILDCB--MRPTKSSSEQSSDAE-----PKPIF--V 128
Db 481 PEIPNGEMLCPRCTCPALKGKQKILIKWGGPPSPTPVPRPPDADPNTSPKAPLGRPE 540
QY 129 KQYLKVKGLSLYHSGSWVPEKE-----FOKAYKSNHRLKTRVNNFHRQME----- 173
Db 541 RQFFVKWQMSYWHCSWVSELQLELHCQVMFNYQRKNDMDPEPPSGDFGDEEKSRRKN 600
QY 174 ----SPNNSDDDF--VAIRPEWTVDRIL--ACREGELELYLVKYKELSDYECWSE--- 224
Db 601 KDPKPAEMEEERFVRYGIKPEWMIHRLNHSVDKKGHVHYLIKWRDLPYDQASWESDVE 660
QY 225 -SDISTFQNEIQRFKDV-----NSRTRRSKQVDHKNPR-----DFQQPDHTBEFLK 270
Db 661 IQDYDLFKQSYWNHRELMRGEGRPGKKLKKYKLRKLERPETPTVPTVKYERQPEYLD 720
QY 271 ---GLLHPYOLGRLNPLRFSWSKOTHVILADBMGLKTIQSIALLASLPEE--NLPHLV 325
Db 721 ATGGTILHPYOMGLNWLRFNSWAQGTDTLADBMGLKTVQTAFLVLSLYKEGHSKGFVLV 780
QY 326 IAPLSTLRNWEREFATWAPQMVNVMYFGTAQARAVIREHEFYLSKDQKKIKKKSGQISS 385
Db 781 SAPLSTIINWEREFENWAPDMYVTVYVGDKSRARIENEFSEFEDNAIRGGKKASRM--- 837
QY 386 ESKQRIKFDVLTLSYEMINLDSAVLKPWKECMIVDEGHLKKNKSKLFSSTIQTSSNH 445
Db 838 -KEASVRFHVLTLSYELITIDMAILGSDWACLIVDEAHLKNNOSKFRVLNGVSLQH 896
QY 446 RILLGTPLQNNLDLFLMLHMFIDAGKFGSLSEFOEFKDIQNEEQISRLHKLAPHLR 505
Db 897 KULLGTPLQNNLEELFLLNFLTPTERFNLHSGFLEEFADIAKEDQIKKLHMLGPHMLR 956
QY 506 RVKQVMDKMPKPKELIIRVDLSSLOKEYYKAFTRNYQVLTKGGA-QISLNNIMWEIR 564
Db 957 RLKADVFKMPSKTELIIVRVELSPMKYKYKYLTRNPEALNARGGQVSLNVVMDLK 1016
QY 565 KYCCHPYMLEGYEPIVHDANEAFKQ-----LLESCGKLQLDKMMVKLKEQGHVRL 615
Db 1017 KCCNHPYLF---PVA--AMEAPKMPNGMYDGSALIRASGKLLLLQKMLNKEGGHRLV 1070
QY 616 IYTOFQHMULDLEDYCTHKKWOYERIDGKVGGAERQIRIDRFNAKSNKFCFLLSTRAGG 675
Db 1071 IFSQMTKMLDLEDFLEHEGYKERYIDGGITGNMQEADIRFNAPGAQQFCFLLSTRAGG 1130
QY 676 LGINLATADTVIYDSDWNPADLQAMARAHRLGQTNKVMYIRLNRGTIEERMMLTKK 735
Db 1131 LGINLATADTVIYDSDWNPADLQAFSAHRIGQNKVMYIRFVTRASVEERITQVAKK 1190
QY 736 KMVLEHLVVGK---LKTQINQBELDDIIRYGSKELFASEDDBAGSKG-----IHYD 785
Db 1191 KMMLTHLVVRPGLSGTSGMSKQELDDILKFGTEELFKDEATDGGDNKEGEDSSVIHYD 1250
QY 786 DAAIDKLLDRDLVBAEVSVDDEENGFLKAFKVFANFEYIDENEAALAEQORVAASKSS 845
Db 1251 DKAIERLLDRNQDETDTEL--QGMNEYLSFSFVAQYVVRSEEMGEERIEIKQES 1308
QY 846 AGNSDRASWEELKDKFELHQAELNALGKRKRKQL-----VSTEEDDLGLAEVDSSD 901
Db 1309 VD-----PDYWEKLLRHHEYOQOEDLARNLGKGRIRKQVNYNDGSDQDRD---WQDDQSD 1361
QY 902 GDSYEAESTDG-EAAGQGGVQTGRPYR---RKGRDNLPTPLMEGEGRSFRVLGNQSQ 957
```


Db 1362 NQSDYSVASBEGDDEPDERSEAPRRPSKGLRNDKOKPLP-PLLARVGGNIEVLGFNARQ 1420
Qy 958 RAIFVQTLMYG-----AGNFDKKEFVRLKQKTFEEINEYGLFLKHIAEEDENSPTF 1012
Db 1421 RKAFLNIMYGMPPQDAFTQW--LVRLRGKSEKEFKAYSLFMHLCPEGADGAET 1478
Qy 1013 SDGVPKEGLRIEDVLRIALLIIVOEKVFVEOHGPKGVFPFPRILERFPGLRSGKIWKEE 1072
Db 1479 ADGVPREGLSRQVLTIGVMSLRKKVQEF----- 1510
Qy 1073 HDKIMIRAVLKHGVRWQAIIVDDKELGIQELIKELNFPHISAAABQAGLOQONGSGS 1132
Db 1511 -----HNGRW-----SMP--ELAEVEENKMSQPGS--P 1536
Qy 1133 NPGAQTNQNGSVITGNNNASADGAOVNSMFFYRDMQRRLVFEVKVRLLEKAMVVEYA 1192
Db 1537 SPKTPSTPGD-----TQNT----- 1553
Qy 1193 EYYGLGSSSIPTEPEAPPKIADTVGVSFIEVDMDLGLPKTDPITSEIMGAAVDN 1252
Db 1554 -----PAPVPAE-----DGIK----- 1565
Qy 1253 NQARVEIAQHYNMCKLLDENARESVOAYVNOPPPKYNESFRALKSINGNI-NTILSI 1311
Db 1566 -----IEN-----SLKEBESIEGEKEVKSTAPETAIEC 1594
Qy 1312 TSDOSKSHEDT-----KPDINNVMKDTAE---ETKPLRGVVVDLNVVEGEENIAE 1360
Db 1595 TQAPAPASEKVVVEPEPEGEKEVKAERVTEEPMETEP--KGAADVKEVE-EKSAID 1651
Qy 1361 ASGSVDVVKMEAEKEEKPKNVV 1383
Db 1652 LTPVIVEDKEKEEKEKVEML 1674

RESULT 4

US-09-976-594-76
; Sequence 76, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR FILING DATE: 2001-10-12
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 76
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 3125723CD1
US-09-976-594-76

Query Match 22.6%; Score 1623.5; DB 4; Length 1739;
Best Local Similarity 30.7%; Pred. No. 1.2e-126;
Matches 453; Conservative 229; Mismatches 493; Indels 299; Gaps 43;
Qy 11 RSDKPPVNLDDSDDDDFVKK-----DRTFEQVEAIVRTDAKENACQACG 56
Db 199 RGKREKQDSDDEDDEDDAPKQTRRAAKNVSYKEDDDPE-----TDS--DDLIEMTG 250
Qy 57 ESTNLVSCNTCTYAFHAKCLVPLPKDASVENWRCEPVSLNEIDKILDCEMRPTKSSEQ 116
Db 251 EGVDEQDNSET-----IEKVID--SRLGKKGAT 277
Qy 117 GSS-----DABPKPIFKV-----QYLVKWKGLSYLHCSWVPEKFEQ-----KAYKS 157

Db 278 GASTTVAIEANGDPSPGDFDTEKDEGEIQYLKWKGWSYIHSTWSEESLQOQVKGLKK 337
Qy 158 NHRLKTRVNFHR-----QWESFNNSDDDFVAJRPETWTVDRILACREE----- 201
Db 338 LENFKKDEIKQWLKGVSPEDVEYFNCOQELASELNKQIQIVVERVIAVKTSTKSTLGQTD 397
Qy 202 -----DGELEYLVKYSDECYWSESDI-STFQNEIORFKOVNSR-----TRR 246
Db 398 FPAHSRKPAPSNEPEYLCKWMLPYSECSWEDEALIGKKFQNCIDSFHSRNNSTKPTIRE 457
Qy 247 SKVDVHRNPRPQQPDHTPEFLKG---LLHPYQLEGLNPLRPSKQTHVILADEMGLG 303
Db 458 CKAL--KORPR-FVALKKQPAVLGGENELRDYQLEGLNWLASHWSCKNNSVILADEMGLG 514
Qy 304 KTIQSTALLASIPFENLI--PHUVIAPLSTRNWEREFATWAPOMVNVVYFGTAQARAVI 361
Db 515 KTIQTISFYUHFHQHLYGFLIVVPLSTLSWQREFEINWPEINNVVITIGLMBRNTI 574
Qy 362 RHEFYLSKDQKKIKKKKSGQISSESKQRIKFPDVLITSYEMINLDSAVLKPICKWECMIV 421
Db 575 REVEMI-----HSQTKLKFNALITTYEILLKDKTVLGSINWAFGV 616
Qy 422 DEGHRUKNDKSLFSSLTQYSSNHRILLTGTPLQNNMLDFMLMFLDAGKPGSLEEFQE 481
Db 617 DEAHRLKNDSDLLYKTLIDFKSNHRLITGTPLQNSLKLWSLLHPIMPEKFEFWDFEE 676
Qy 482 EPKDINQEQISRLHKLAPHLRLRVKDKVMDMPKELILRLVDSLSSLQKEYYKAIATR 541
Db 677 DHGK-GRENGYQSLHKLVEPFLRLRVKDKVEKSLPAKVQILRVMSALQKQYKWLTR 735
Qy 542 NYQVLT-KGGAQISLNNIMMELRVCCHPYMLGEVPE-VIHDANAPKQLLESCKGLQL 599
Db 736 NYKALAKGTGSTSGFLNIVMLKCCNHCYLKPPENERENGQBIILLSSGKLIL 795
Qy 600 LDKQWVKLKEQHRVLIYTFQHMILLDLYCHYKQWYERIDGKVGGAERQIRIDFNA 659
Db 796 LDKLLTLRERGNRVLFQSMVRLDILAEYLTIKHYFPQRLDGSIGKIRKQALDHFA 855
Qy 660 KSNKFCFLSTRAGGLINLATADTVIYDSDNPHADLOAMARAHRLGQTNKVMYRL 719
Db 856 DGSDFCLLSTRAGGLINLASADTVIFDSDNPNQDLQAQARAHRLGQKQVNIYRL 915
Qy 720 INRGTIERMMQTKKRWMLHLV-----GKLTQNIQOEELDDIIRYG 764
Db 916 VTKGTVEEBIIBRAKKKQWLDHLVIQRMDTGTGTILENNSGRSNPNFNKEELTAILKFG 975
Qy 765 SKELFASDEDEAGSKIIHYDDAAIDKLLDRDLVEAEVSVDDDEENGFLKAFKANFEY 824
Db 976 AEDLFKELEGESEEP-----QEMDIDEIL--RLAETRENEVSTSATDELLSQPKVANFAT 1028
Qy 825 IDENEAALAEQRAVAAESKSSAGNSDRASVWEELL-----KDKFELHQAEBELNALGK 876
Db 1029 MEDEE--ELEER-----PHKD-----WDBIIEBORKKVEEERQKELEIYMLPR 1072
Qy 877 RKESRKLVSIEDDLIAGLEDVSSDGDSEYAE-STDGEAAGQGVQTRRPRYRKGRDNL 935
Db 1073 IRSSTKK-----AQTNDSDDTESKQQAQSSASESETEDSDDKPKER----- 1117
Qy 936 EPTPLMEGRGRSPR-----VLGFNQSORAIFVQTLRMRYGAGNFDWKEFVPR---LKQKTFEE 989
Db 1118 -----GRPSRVKDLVEGFTDAEIRRFKAYKKFGL-PLERLECIARDAELVDSKVD 1169
Qy 990 INEYGTILF-----LKHTAEEDENSPTFSDGVKPEGLRIEDVLVRIALLIIVQEKVKF 1042
Db 1170 LKRLGELIHNCSVAMQYEEQKENASGKGPCKRRGPTIKISGVQVNVKSIHQHEEF 1229
Qy 1043 VEDHPGKVPFPPSR-----ILERFPGLRSGKIWKEEDKIMIRAVLKHGVRWQAIIVDDKE 1097
Db 1230 EMLHKSIPVDPEBKVKYCLTRVKAHFVDEWVEDDSRLLLGIYEHGYGNWELIKTDPE 1289
Qy 1098 LGTQELIC-----KELNPPHLSAAEQAGLOGQSGSGSNPAGQTNQPSVIT 1147

Db 1290 LKLTDKLPVETDKKQKQLOTRADYLLKLLRKGLEKK-----GAVT 1332
Qy 1148 GNNASADGAQVNSMFYRDMQRLVFEVFKRVLLLEKAMNBYAEYVGLGSSSIPT 1207
Db 1333 GGBEAK-----LKKRPVKENKVPRLKEEHGIELSSPRHS 1370
Qy 1208 EPEAEPIADTVGVSFIEVDDEMLDGLPKTDPITSSEINGAAVDNNQARVEIAQYNMC 1267
Db 1371 NPSEGEVKD-----DGLEKS-PMKKKQKKENKENKE-----KQMS 1406
Qy 1268 KLDD-ENARESVOQVYNNQPPSTKVNESFRALKSINGNINTILSITSDQSKSHEDDTKPD 1326
Db 1407 SRDKEGDKERKSKDKKPKSGDAKSSKRSQGPVHITAGSPVPVIGEDDD---D 1463
Qy 1327 LNNVEMKOTABETKPLRGVVDLNVVEGENIAE 1360
Db 1464 LQGETSICKERMVPVKALKOLDKPKGLNVQE 1497

RESULT 5
US-09-535-008-67
; Sequence 67, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigan, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; TITLE OF INVENTION: AND OTHER CANCER TYPES
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535, 008
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 67
; LENGTH: 1646
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-67

Query Match 14.3%; Score 1028.5; DB 4; Length 1646;
Best Local Similarity 30.8%; Pred. No. 1e-76;
Matches 300; Conservative 156; Mismatches 324; Indels 195; Gaps 32;

Qy 269 LKGLLHPYQLEGLNFRFWSKQTHVILADEMGLGTIQSIALLASIFRENLI--PHLVI 326
Db 750 VNGVLKQYQIKGLEWLVSLYNNNLGILADEMGLGTIQTIATITYLMEHKLINGFLII 809

Qy 327 APLSTLNNWEREPATWAPQNVVYFQTAAQARAVIREHEFYLSKQKKIKKSGQISSE 386
Db 810 VPLSTLSNWAYEFDKAPSVVYKSGSPAARRAF-----VPQLRSG----- 851

Qy 387 SKOKRIKFDVLLTSYEMINLDSAVLPKWECHIVDEGHLKNKDKLPSLL-TQYSSNH 445
Db 852 -----KFNVLTYTYEIIKDKHLAKIRWKYIVDEGHRMKNHCKLTQVLNTHYVAPR 905

Qy 446 RLLTGTPLONNDELFLMHFLDAGKFGSLEREFQEFK-----DINQEQ--I 492
Db 906 RLLTGTPLQNKLPFWALLNLLPITFKSCSTFEQWFAFNAPPANTGEKVDLNEEETILII 965

Qy 493 SRLHKMLAPHLRRVKDVKMPPKKELILRDLSSLOKEYY-----KAIFTRNYQVLT 547
Db 966 RRLHKVLRPFLRLRLKEVEAQLPEKVEYVVKDMSALQVLYRHMQAKGVLLTDGSEKD 1025

Qy 548 KKG-GAQISLNNIMELRKCCHPMLGEVPEVHIDANEAPKO-----LLES 593
Db 1026 KKGKGGTKTLNMTIMQLRKICNHPYMFQHIE-----ESFSEHLGFTGIVQGLDLYRA 1078

Qy 594 CGKLOLDDKMWVKLEQGHVLYTQFQHMLDLLEDYCTHKWQYERIDCKVGGAERQIR 653
Db 1079 SGKPELDLRLPKLRATNKHVLLFCQMTSLMTIMEDYFAYRGPKYLRLDGTTKAEORGLM 1138

Qy 654 IDRFNAKSNKFCPLLSTRAGGLINLATADTVIYDSDWNPHADLOAMARHLQOTNK 713
Db 1139 LKTFEPGSEGYFIFLLSTRAGGLNLQADTVIIFDSDWNPQDLOAQDRAHRIQOQNE 1198

Qy 714 VMYRLINRGTTIERRMQLTKKQWVLEHLV--GKLKTONIQOE--LDDIIRYSGKEL 768
Db 1199 VRVLRCTVNSVEEKLAAAKYKLVNDQKVQAGFDDQSSSHERRAFQAILEH----- 1253

Qy 769 FASEDE-----AGKSGKIHY--DDAAIDKLLDRDLVEAEVSVDDDEENGFLKAFKV 819
Db 1254 --EEDESRHCTGSGSASFATPPAGVNPDPLEBPPLKEDEVPDDTVNQMIARHE- 1310

Qy 820 ANFEYIDENEAALAEARVAES-----KSSAGNSDRASYWEELLKDFELH-----QAEELN 872
Db 1311 ---BEFDLFMRDLDRRREEARPKRKLMBEDELPSW--IHKDAEVRERLTCEBEEK 1365

Qy 873 ALGKRKRSRKQ-----LVSIEDDDLAGLEDV-----SSDGDDESIEA 908
Db 1366 MFGGRSRHRKEVDYSDSLTEKQWLKAIEBGTLEETEEVEVRQKSSRRKRKDSAGSSTPT 1425

Qy 909 EST-----DGEAAGQGVQTRRPYRRKGR---DNLEPTP-----LWEGE 944
Db 1426 TSTRSRDKDDESKQ-----KCRGRPPAKKLSPPNPPLTKMKKIYDAVIKYKDS 1476

Qy 945 GRSFVLFQNSQRAIFVQTLRMRYGAGNF-----DWKEFVPRLKQKTFEINEYGIL 996
Db 1477 GRQL-----SEVFIQLPSRKELPEYELIRKPVDFKTKERINNHKYSLND--- 1523

Qy 997 FLKHTAEIEDENSPFSDGVPKEG-LRIEDVLVRIALLIVOEKVFVEDHPGKVPFPSR 1055
Db 1524 -LEKDVMLCQNAQTFN-----LEGSLIYEDSVLQSVFTSVRQKIEKEDDSEGE-----E 1573

Qy 1056 ILERFPGLRSKGIWKEBHDKIMIRAVLKHGYGRWQAIYDDKELIGIQLICKELNPHISL 1115
Db 1574 SEEEEBEGEESSESRSVKVVK-----LGRKEKAQDLKGRRRRPSRGSRAKPVWSD 1627

Qy 1116 SAAEQAGLOGQSG 1130
Db 1628 DDSEEBEEDRSGSG 1642

RESULT 6

US-09-535-008-2
; Sequence 2, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigan, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; TITLE OF INVENTION: AND OTHER CANCER TYPES
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535, 008
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1647
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-2

Query Match 14.3%; Score 1027; DB 4; Length 1647;
Best Local Similarity 30.9%; Pred. No. 1.3e-76;
Matches 301; Conservative 157; Mismatches 325; Indels 190; Gaps 32;

Qy 269 LKGLLHPYQLEGLNFRFWSKQTHVILADEMGLGTIQSIALLASLFFENLI--PHLVI 326
Db 750 VNGVLKQYQIKGLEWLVSLYNNNLGILADEMGLGTIQTIATITYLMEHKLINGFLII 809

Db 1477 DSSGRQL-----SEVPIQLPSRKELPEYVELIRKPVDFKKIKERIRNHKYSRLND- 1526
Qy 994 GILFLKHIAEIDENSPTSDGVPKGG-LRIEDVLRIALLILVQSKVPVEDHPKQPVF 1052
Db 1527 -----LEKVMLLCQNAQTFN-----LEGSLEYEDSIVLQSVFTSVRKIEKEDDSEGE--- 1575
Qy 1053 PSRLERFPGLRSKGIWKEEHDKIMIRAVLKHGCGYQWQAIIVDDKELGIGELICKELNFPFH 1112
Db 1576 --ESEEEEGEESGESRSVKIK-----LGRKEAQRDLKGRRRRPSGRSAKPV 1627
Qy 1113 ISLSAAEQAGLQGGSG 1130
Db 1628 VSDDEEERDEDRSGG 1645
RESULT 8
US-09-535-008-71
; Sequence 71, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigan, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; TITLE OF INVENTION: AND OTHER CANCER TYPES
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535, 008
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 71
; LENGTH: 1650
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-71

Query Match 14.3%; Score 1025.5; DB 4; Length 1650;
Best Local Similarity 30.8%; Pred. No. 1.8e-76;
Matches 301; Conservative 157; Mismatches 325; Indels 193; Gaps 32;

Qy 269 LKGLHPYQLEGFLRFSWQTHVILADEMGLGKTIQSIALLASLFEENLI--PHLVI 326
Db 750 VNGVLKQYQIKGLEWLVSLYNNLNLGILADEMGLGKTIQTIALITYLMEHKKRNGPFLII 809
Qy 327 APLSTLRNWEREFATWAPQMVVMYFCTAQARAVIREHEFEYLSKQDKIKKKSGQISSE 386
Db 810 VPLSTLSNWAYEFDKWPVSVKYSKGPAAARAF-----VPQLRSG----- 851
Qy 387 SKQRIKFDVLLTSYEMINLDSAVLKPWKECMIVDEGHLRKNKOSKLFSSL-TOYSSNH 445
Db 852 -----KFNVLTTYEVIKDKHILAKIRWKYMIIVDEGHRMKNHCKLTQVLNTHYVAPR 905
Qy 446 RILLTGTPLQNNLDELPMHFLDAGKFSLEFQEEFK-----DINQSEQ---I 492
Db 906 RLLTGTPLQNKLPPELWALLNPLFTIFKSCSTFEQMFNAPPAMTGEKVDLNEEETILII 965
Qy 493 SRLHKMLAHLRRVRKVDYMKMPKELIRVDLSLQKEYY-----KAIFTRNYQVLT 547
Db 966 RLHKVLRPFLRLKKEVEAQLPEKVEYVVKOMGALORVLYRHMQAQGVLLTDGSEKD 1025
Qy 548 KKG-GAQISLNNIMELRVCCHPYMLGEPVPIHDANEAFKQ-----LLSES 593
Db 1026 KKGKGGTKTLNMTIMQLRKICNHPYMFQHLE-----ESFSEHGLGFTGGIVQGLDLYA 1078
Qy 594 CGKLQLLDKMVKLKEQGRVLIYTFQFHMLEDLECYTHKKWQYERIDGKVGGAERQIR 653
Db 1079 SGRFELLDRILPKLRATNHNKVLFCQMTSLMTIMEDYFAYRGFKYLRDLGTTKAEDRGM 1138
Qy 654 IDRFNAKSNKFCFLSTRAGGIGINLATDVIIYDSWNPHADIQAMARHLGQTNK 713
Db 1139 LKTFNPGSEYFILLSTRAGGIGLNLSQADTVIIFDSWNPHQDLQAQDRAHRIQQONE 1198

Qy 714 VMIYRLINRGTTIERMMQLTTKKMWLEHLV--GKLKTONIQEE---LDDIIRYGSSEL 768
Db 1199 VVRLRLCTVNSVEEKILAAAKYKLANVDQKVIQAGMPDQKSSSHERRAFQAILEH----- 1253
Qy 769 FASEDE-----AGKSGKIHV--DDAAIDKLLDLDLVEAEVSVDDDEENGFLKAFKV 819
Db 1254 --EQDESCHCTSGSASFATAPPGVGNPDLEPPLKEDEVEDDDTVNQMIARHE- 1310
Qy 820 ANFEYIDENEAALSAQRVAAES---KSSAGNSDRASYWEELLKOKFELH---QAEELN 872
Db 1311 ---EEFDLPMRMDLDRRREEARNPKRPLMEDEDELPSW--IIKDDAEVERLTCEBEEK 1365
Qy 873 ALGKRRSRKQ-----LVSIEDDLAGLEDV-----SSGDSES 905
Db 1366 MFGGRSRHRKEVDYSDSLTEKQWLKTLKAIEBETLEIEEVEVRQKSSRRKRDSDAGSS 1425
Qy 906 YEAEET-----DGEAAGQGVQTRRPYRRKGR--DNLEPTP-----LMEG 943
Db 1426 TPTTSTRDKDDESKQ-----KKGRPPAEKLSFNPPLTKMKKIVDAVIKYK 1476
Qy 944 EGRSPVLGFGNQSORAIFVQTLMYGAGNF-----DWKEFVPLKQKTFEINEYGI 995
Db 1477 DSSSGRQLS-----EVFIQLPSRKELPEYVELIRKPVDFKKIKERIRNHKYSRLND--- 1527
Qy 996 LFLKHIAEIDENSPTSDGVPKGG-LRIEDVLRIALLILVQSKVPVEDHPKQPVFP 1054
Db 1528 --LEKVMLLCQNAQTFN-----LEGSLEYEDSIVLQSVFTSVRKIEKEDDSEGE----- 1576
Qy 1055 RILERFPGLRSKGIWKEEHDKIMIRAVLKHGCGYQWQAIIVDDKELGIGELICKELNFPFH 1114
Db 1577 ESEEEEGEESGESRSVKIK-----LGRKEAQRDLKGRRRRPSGRSAKPVVS 1630
Qy 1115 LSAAEQAGLQGGSG 1130
Db 1631 DSDSEEEQEDRSGG 1646

RESULT 9

US-09-535-008-77
; Sequence 77, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavtigan, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRG1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; TITLE OF INVENTION: AND OTHER CANCER TYPES
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 77
; LENGTH: 1681
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-535-008-77

Query Match 14.1%; Score 1011; DB 4; Length 1681;
Best Local Similarity 29.7%; Pred. No. 3.1e-75;
Matches 300; Conservative 156; Mismatches 324; Indels 230; Gaps 32;

Qy 269 LKGLHPYQLEGFLRFSWQTHVILADEMGLGKTIQSIALLASLFEENLI--PHLVI 326
Db 750 VNGVLKQYQIKGLEWLVSLYNNLNLGILADEMGLGKTIQTIALITYLMEHKKRNGPFLII 809
Qy 327 APLSTLRNWEREFATWAPQMVVMYFCTAQARAVIREHEFEYLSKQDKIKKKSGQISSE 386
Db 810 VPLSTLSNWAYEFDKWPVSVKYSKGPAAARAF-----VPQLRSG----- 851

Qy 880 SRKQVSI EEDDLA GLED 897
| | : : | | |
Db 630 KRXVLSPEB-----LED 642

Search completed: September 14, 2004, 02:09:21
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 14, 2004, 02:05:02 ; Search time 88 Seconds
(without alignments)
5043.573 Million cell updates/sec

Title: US-10-049-137-2

Perfect score: 7187

Sequence: 1 MSSLVERLRIRSDRKPVNL.....VDVKMEAEKBEKPKMVD 1384

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA.*
- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/2/pubpaa/PCTU5_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
 - 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3752	52.2	1358	16	US-10-675-072A-43
2	3621	50.4	1364	16	US-10-437-963-155910
3	2105.5	29.3	1954	12	US-10-462-261-2
4	2083	29.0	1944	16	US-10-408-765A-545
5	1874.5	26.1	1979	15	US-10-104-047-2446
6	1621.5	22.6	1739	15	US-10-341-434-61
7	1570.5	21.9	1795	8	US-08-973-363-17
8	1555	21.6	2108	8	US-08-973-363-16
9	1551.5	21.6	1434	8	US-08-973-363-20
10	1548.5	21.5	2192	16	US-10-437-963-201156
11	1536.5	21.4	1465	15	US-10-369-493-5100
12	1492.5	20.8	1967	10	US-09-849-602-16
13	1473	20.5	1407	15	US-10-369-493-2310
14	1461.5	20.3	1467	8	US-08-973-363-21
15	1461.5	20.3	1468	15	US-10-369-493-1775

15	1426.5	19.8	1264	15	US-10-369-493-3742	Sequence 3742, Ap
16	1407.5	19.6	2131	16	US-10-408-765A-1179	Sequence 1179, Ap
17	1366	19.0	1413	16	US-10-437-963-138262	Sequence 138262, A
18	1356	18.9	1061	15	US-10-369-493-12847	Sequence 12847, A
19	1293.5	18.0	1141	16	US-10-437-963-106873	Sequence 106873, A
20	1241	17.3	1152	12	US-10-412-699B-1581	Sequence 1581, Ap
21	1241	17.3	1158	16	US-10-437-963-154461	Sequence 154461, Ap
22	1241	17.3	1158	16	US-10-437-963-154461	Sequence 154461, Ap
23	1230	17.1	1104	12	US-10-425-114-57274	Sequence 57274, A
24	1224.5	17.0	1057	10	US-09-934-455-432	Sequence 432, App
25	1224.5	17.0	1057	12	US-10-412-699B-790	Sequence 790, App
26	1224.5	17.0	1057	12	US-10-412-699B-1846	Sequence 1846, Ap
27	1224.5	17.0	1057	12	US-10-225-066A-1044	Sequence 1044, Ap
28	1224.5	17.0	1057	15	US-10-374-780A-2554	Sequence 2554, Ap
29	1215	16.9	1122	12	US-10-412-699B-1582	Sequence 1582, Ap
30	1215	16.9	1122	16	US-10-437-963-151729	Sequence 151729, A
31	1157	16.1	730	16	US-10-408-765A-1039	Sequence 1039, Ap
32	1147	16.0	659	12	US-10-424-599-226981	Sequence 226981, A
33	1142	15.9	620	12	US-10-425-114-37386	Sequence 37386, A
34	1139	15.8	1120	15	US-10-369-493-22416	Sequence 22416, A
35	1137.5	15.8	956	15	US-10-369-493-3306	Sequence 3306, Ap
36	1135.5	15.8	1143	15	US-10-369-493-1499	Sequence 1499, Ap
37	1131	15.7	1027	14	US-10-205-219-123	Sequence 123, App
38	1126	15.7	971	15	US-10-369-493-5755	Sequence 5755, Ap
39	1120	15.6	1359	15	US-10-369-493-22090	Sequence 22090, A
40	1097	15.3	1302	14	US-10-032-585-7534	Sequence 7534, Ap
41	1095.5	15.2	1703	9	US-09-801-368-340	Sequence 340, App
42	1095.5	15.2	1703	10	US-09-824-574-3	Sequence 3, Appl
43	1095.5	15.2	1703	15	US-10-369-493-22413	Sequence 22413, A
44	1093.5	15.2	1474	15	US-10-369-493-6164	Sequence 6164, Ap
45	1093.5	15.2	1474	15	US-10-369-493-6165	Sequence 6165, Ap

ALIGNMENTS

RESULT 1

US-10-675-072A-43
; Sequence 43, Application US/10675072A
; Publication No. US20040098760A1
; GENERAL INFORMATION:
; APPLICANT: Yumin, Tao
; APPLICANT: Gordon-Kamm, William
; APPLICANT: Shen, Bo
; APPLICANT: Lowe, Keith
; APPLICANT: Danilevskaya, Olga
; APPLICANT: Mahajan, Pramod
; APPLICANT: Rafalski, Jan Antoni
; APPLICANT: Sakai, Hajime
; APPLICANT: Klein, Theodore
; TITLE OF INVENTION: Transcriptional Regulatory Nucleic
; FILE OF INVENTION: Acids, Polypeptides, and Methods of Use Thereof
; FILE REFERENCE: 1289
; CURRENT APPLICATION NUMBER: US/10/675,072A
; CURRENT FILING DATE: 2003-09-30
; PRIOR APPLICATION NUMBER: 10/005,057
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/251,555
; PRIOR FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 1358
; TYPE: PRT
; ORGANISM: Zea mays
US-10-675-072A-43

Query Match 52.2%; Score 3752; DB 16; Length 1358;

Best Local Similarity 54.7%; Pred. No. 1.5e-272;

Matches 765; Conservative 227; Mismatches 305; Indels 102; Gaps 32;

QY 1 MSSLVERLRIRSDRKPVNLDDDDDFVP-----KKDRTFEQVEAIVRTAKENAC 52

Db 1 MSSLVERLRIRSDRKPVNLDDDDDFVP-----KKDRTFEQVEAIVRTAKENAC 58

Qy	53	QACGESTNLVSCNTCTVAFHAKFLVPLPKDASVENWRCPECVSPLEINBDKILDCMRPTK	112
Db	59	QRCGKSNLVSCSTCTYKFKRKLVLPCNLITS-DKWSCEPCVSPLTMYERILDIEV--LE	115
Qy	113	SSQGSDDAPKPIFVKQYILVKWKGLSYLHCSWVPEKFOKAKYSHRLKTRVNVNFRQM	172
Db	116	APRSDSSTPRSGKMMERYILKWKGLSYIHCSWVSEKEYGEAANIHPRLTRLNFRQK	175
Qy	173	ESF---NNEDDFVALRPEWTTVDTRILACREE-DGELEVLVKYKELSYDECYWESES	227
Db	176	EAWKIEAERGEDIVALRPEWTTVDTRLASRKNLVGREYIVKWNELTYEECTWENES	235
Qy	228	STFQNEIQRFDKVNSTRRSKVDHKNRPDFOQDHTPEFLK-GLLHPYQLQEGLNFLRF	286
Db	236	TVQPEIERFNEI--QPRRKGSGDKATREPQFKESPFLSGGLTLPYQLQEGLNFLRY	293
Qy	287	SWSKQTHVILIADENGLGKTTOSIALLASLPEENILPHLVTAPISTLXNWEREFATWAPQ	346
Db	294	SWFHNRKVLGDENGLGKTTQSIATLASLFPDGHPLVVAPLSTLXNWEREFATWAPQ	353
Qy	347	NVMYFCTAARAVIREHEFVLSKQ--KTKKKXSGOISSESQKRITKFDVLLTSYEMIN	405
Db	354	NVMYFGAASRDIIIRKEYFYYPKEUKULKKKXSSPNSDEKKQSRIRFDVLLTSYEMIN	413
Qy	406	LDSAVLKPIKWECMIVDEGHRLKNKDSKLFSSLTQYSSNHRILLTGTPLQNNLDELFLML	465
Db	414	MDSAILKNIWECLVDEGHRLKNKDSKLFQOLKYNTKURVLLTGTPTVQNNLDELFLML	473
Qy	466	HFLDAGKFGSLBEFQEBFKDINOBEQISRLHKMLAPHLLRRVKDKVMKMPKPKELILRV	525
Db	474	HFLGESFGSITDLOEBFKDINOQKIEKLHGMLKPHLLRRFKDKVMKELPKPKELILRV	533
Qy	526	DLSSLQKEYYKATPRNYQVILTKKGAQISLNNIMELRKVCCHPYMLGEVPEVIHDAN-	584
Db	534	ELTRKOKEYYKAILTKNYEVLARRNGHTSLINVMEELRKLCCGHFMID--EPDLEPANP	591
Qy	585	EAFKOLLESCKGLQLLDKMMVKLKEQCHRVLIYTOFHMLDILEDCYTHKKWQYERIDG	643
Db	592	EEGLRLLDSGKMWQLLDKMMVKLKEQCHRVLIYQOQHMLDILEDCYLSYRKWTYERIDG	651
Qy	644	KVGAERQIIRIDRPNAXNSKFCFLSTRAGGLGINLATADTVIIYSDSNWPNHADLOAMA	703
Db	652	KISGADRQIIRIDRPNAXNSRFCFLSTRAGGLGINLATADTVIIYSDSNWPNHADLOAMA	711
Qy	704	RAHRLGQTKVMYRLINRGTIETERRMWOLTKKXVLEHVLVVKL-KTONINOELDDIIR	762
Db	712	RAHRLGQTSKVMYRLVSRGTIETERRMWOLTCKKTLLEHLVVGRLTKANNVNOELDDIIR	771
Qy	763	YGSKELFASDEDEAGSGKTHYDDAAIDKLLDRDLVEAEVSVDDREENGFLKAFKXVNF	822
Db	772	YGSKELFEDENDE---SRQIHYDPAALERLLDRDQVDGDB-SVEDEEEDGFLKGFKVNF	827
Qy	823	EYIDENEAFA--LEAQRVAAESKSSAGNSDRASWHEELKDKFELHQAELNALGKGRS	880
Db	828	EYIDEAKAQAEKEARRKAA--AEAEENSER-NYWDPELLKDRYDVQVQVEHTAMGKGRS	883
Qy	881	RKQLVSTIETHDDLAGELVSDGDBESYAEASTDGBAAGQVOTGHR-PYRRKGRDNLEPTP	939
Db	884	RKQMAADEDDIHDLS--SEDEYSLEDDISDNDTSLQGNISGKRGQYSRKRSRNVDSIP	941
Qy	940	LMEGEGRSFVLGNQSORAIFVOTLMRYGAGNFDWKEFVPLRKQKTFEINEYGIILFLK	999
Db	942	LMEGEGRTLVLGNFNAQRAFLQTLNRFQONTDWMKEYIFLRLKGKSVSEIORVAELVMA	1001
Qy	1000	HIABEIDENSPFSDGVPKBEGLRIEDVLVRIALLVQEKVKFVEDHPGK--PVFPSSRIL	1057
Db	1002	HLVEEIND-SDYFSDGVPKEMRWVDDVLRIANISLIEEKW--AATGFKITNIFPNYLL	1058
Qy	1058	ERFPGLRSGKITKEEHDKIMIRAVLKHGYGRWQAIVDDKELGIQELICKENLFPNHSLSA	1117
Db	1059	YEFQGLSGGRWKAHEHLLLLRILGILKHEGYARWQVYISDDRENGLPEAARREHLHPVSN--	1115

RESULT 2

```

US-10-437-963-155910
US-10-437-963-155910
; Sequence 155910, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 155910
; LENGTH: 1364
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_55629C.1.pgp
US-10-437-963-155910

```

[illegible]


```
Qy 557 NNIMMELKVCCHPYM--LEGVE-PVITHDANEAFKQLLESCGLQLLDKMMVKLEQGH 613
Db 983 LNIWMDLKKCNHYPFPVAAVEAPVLPNGSVSSLVKSGKMLLQKMLKKLRDSGHR 1042
Qy 614 VLIYTOQHMLDLECYTHKKQYERIDGKVGAEQRIIDRPNKSNKFKFELLSTRA 673
Db 1043 VLIFSOMTKMLDLECYTHKKQYERIDGKVGAEQRIIDRPNKSNKFKFELLSTRA 1102
Qy 674 GGLGINLATADTVIYDSNPNHADIQAFRAHRIQGNKKVMYRFRVTRASVEERITQVA 733
Db 1103 GGLGINLATADTVIYDSNPNHADIQAFRAHRIQGNKKVMYRFRVTRASVEERITQVA 1162
Qy 734 KKKMVLHVLVVGK---LKTQINQEELEDDIIRYKSKELFASE-----DD 774
Db 1163 KKKMVLHVLVVGK---LKTQINQEELEDDIIRYKSKELFASE-----DD 1222
Qy 775 EAGSGK-----HYDDAAITDKLLDRPLVEAEVSVDDE 809
Db 1223 QSSKGGNLAASAKKKHGSTPPGDKOVDSVHYDDAAITDKLLDRN-----QDATDDTE 1277
Qy 810 ---ENGLKAFKVFANFEYIDENEAALAEQVAEAKSSAGNSDRASYWEELKDKFELH 866
Db 1278 LQNNNEYSFKAQYVVRBEDGVVEVEREIIKQEE-----NVD-PDYWEKLLRHYYEQ 1331
Qy 867 QAEELNALGKRKRKQKQ---VSIEDDLAGLEDVSSDGDESY-----EAESTDGEAAQ 918
Db 1332 QEDLARNLGGKRIKQVNYNDASQEDQE---WQDELSDNQSEYSGSDEDEDFEERPE 1388
Qy 919 GVQGRPRPYRK---GRDNLEPTLMGEGRSFRVLGNOSQRAIFVQTLMRVG-----A 970
Db 1389 G-OSGRQRRLKSDRDKPLP-PLLARVGNIIEVLGNARQKAFNALMRWGMPPQDA 1446
Qy 971 GNFDKQFVRLKQKTEEINEYGLFLKHAIEIDENSPTESGVPKEGLRIEDVLVRI 1030
Db 1447 FNSHW---LVRLDRGKSEKFRAYVSLFMRHLCEPGADGAETFDAGVPRGELSGHVLTRI 1504
Qy 1031 ALILVQKVKFVBDHPCKVFPFSRILRFPGLRSRKIMKEEHDKIMIRAVLKHGYGRWQ 1090
Db 1505 GVMVLKVKVQEFHVGKYSTPDLIPEGPEGKSGEV-----1542
Qy 1091 AIVDDKELGI---QELIKELNPPHISLSAAEQAGLOQNGSGGSPGAGTQNPQSVI 1146
Db 1543 -ISSDPNTPVPASPAHLPLAPLGLPD---RMEAQLGYMDE---KDPGAKPRQP---1589
Qy 1147 TGNNASADGAQVNSMFYRDMQRRLVEFVKRVLILLEKAMNVEYAEYVGLGSSSIPT 1206
Db 1590 -----LEVOALPAALDRVSEDKH-----1608
Qy 1207 EPEAPEKIAIDTVGVSFIEVDDMLDGLPKTDPTITSEEIMGAADVNNQARVEIAQHYNQ 1266
Db 1609 ESPASKERARE-----ERPEETEKAAPPSPQPLPREVL-----PEK 1644
Qy 1267 CKLLDENARVSQAYVN---QPSTKNSFRALKSINGNINTILSITSDQSKSHDDT 1323
Db 1645 EKILDKLESLIHSRGSSSELRLPDPTKAEKEPIETQONGD-----KEEDDEG 1692
Qy 1324 KPLDNNVEMK---DTAETKPLRGVVDLVNV-RGEENIARASGSV 1365
Db 1693 KKEDKKGKFKPMFNIAD-----GGFTELTLWQNEERAUVSSGI 1732
```

RESULT 4

```
US-10-408-765A-545
; Sequence 545, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
```

```
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408, 765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 545
; LENGTH: 1944
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-408-765A-545
```

```
Query Match 29.0%; Score 2083; DB 16; Length 1944;
Best Local Similarity 34.3%; Pred. No. 9.7e-147;
Matches 520; Conservative 211; Mismatches 415; Indels 370; Gaps 41;
```

```
Qy 52 QACGSGSTNLVSNCTCYAFHAKLVPPPLKADSVENWRPECVSPPL--NEIDKILDCMR 109
Db 459 CRVCKDGGELLCCDACITSSYHICLNPPPLDIPNGEWLCPRCTCPVLKGRVQKILHWRWG 518
Qy 110 -----PTKSSSEOGSSDA-EPKPI---FVKQVLVKKGLSYLHCSWVPEKEFOKAYKSNH 159
Db 519 EPPVAVPAQOQADGNPDVPPRPLQGRSEREFVKKWGLSYWECSSWAKELQLIFHLVMY 578
Qy 160 RLKTRVNNFHRQME-SFNNSEDD-----FVAIRPEWTTVDRI-- 196
Db 579 RNYQRKNDMDPEPPLDYGSGEDDGSKRKKVDPHYAEMEKKYRFGIKPEWMTVHRIIN 638
Qy 197 ACREEGELEYLVKYLKSYDECYW-ESSDSDISTFQNEIQ---RFKDVNSRTRRSKDVHD 252
Db 639 HSDVKKGNVHYLVKWRDLPYDQSTBEEDENMIPEYBEHKQSYWRHRELINGEDPAQPRKY 698
Qy 253 KRNPDRFQ-----QFDHTPRFLK---GLLHPYQLEGLNLFKFSKOTHVI 295
Db 699 KKKKELQGGPPSSPTNDPTVKYETQPRPTATGTTLHMVYQLEGLNLFKFSWAQGTDTI 758
Qy 296 LADEMGLGKTIQSIALLASLFEENLI--PHLVIAPLSTLRNWEREFATWAPQNNVVMYFG 353
Db 759 LADEMGLGKTIQITVFLYSLYKSGHTKGPPLVSGAPLSTIINWEREFQMWAPKVVVYTG 818
Qy 354 TAQARAVIREHPEYLS-----KQOKKIKKKSGQISSEKQRIKFDVLLTSYEMINLDSA 409
Db 819 DKDSRAIIRENEFPEDNAIKGKKAFOK-----REAQVFFHVLTSYELITDQA 870
Qy 410 VLKPIKWECMIVDEGRLKNDKSLFSSLTQYSSNHRILLTGTPLQNNLDELFLMHLFD 469
Db 871 ALGSIKWACLVVDEAHLKNNQSKFRVLNGYKIDHKLLLTGTPLQNNLELPHLNLFLT 930
Qy 470 AGKFGSLSEEFQEEFKDINOBEQISRLHKMLAPHLLRRVKDKVMKMPKPKKELILRDLSS 529
Db 931 PERFNLEGLFEADISKEDQIKKLDLGLPHMLRLKADVPKNMPAKTELIVRVLESP 990
Qy 530 LQKEYYKAIPTRNYQVLTCKGGA-QISLNNIMMELKVCCHPYMSEGVEVIHDANEAFK 588
Db 991 MQKKYKYILTRNFEALNSRGNGQVSLNIMMDLKKCCNHPYLF-----EVA--AMESPK 1044
Qy 589 -----QLLESCGKLQLLDKMMVKLEQGHRLVLIYTOFHMLDLECYTHKKQYE 639
Db 1045 LPSGVEGGALIKSSGKMLLQKMLKLEQGHRLVLIYTOFHMLDLECYTHKKQYE 1104
Qy 640 RIDGKVGGAERQIRIDRFNAKNSNKFCLLSTRAGGLGINLATADTVIYDSNPNHADI 699
Db 1105 RIDGITGALRQEAIDRFNAPGAQCFLLSTRAGGLGINLATADTVIIFDSNPNHADI 1164
Qy 700 QAMARAHRIQTNKVMYRLINRGTTTEERMMQTLTKKQWLEHLVVGK---LKTQINQEE 756
Db 1165 QAFSRAHRIQANKVMYRFRVTRASVEERITQVAKRQOMLTHLVVRPGLSGKAGSMKQ 1224
Qy 757 LDDIIRYKSKELFASDEDEAGK---SGKTHYDDAAITDKLLDRPLVEAEVSVDDEENG 813
Db 1225 LDDILKPGTEELFDENEGENKEEDSSVIHYDNEATRLDRNODATEDTDTV--QNNNEY 1282
```

Qy	814	LKAPKVANPEYIDNEAAALAEAQRAVAASKSSAGNSDRASYWELLKKDKFELHQAEBLNA	873
Dy	1283	LSSFKAQVYVVREBDKIEETIEREIIKQEE-----NVD-PDYWEKLHRHHVEQQOEDLAR	1336
Qy	874	LGEKEKRSKQL---VSIEEDDLAGLENDSDGDESYEAESTDGEAAGQGVOTGRPPYRK	930
Dy	1337	LGGKRVKOVNYNDAQED-----QNQSSEYSVGSEEDDEFDERPEGHRQSKRQLRNE	1391
Qy	931	GRDNLEPTPLMEGEGRSFRVLGFNQSORAIFVOTLMRYG-----AGNFWDKFEVPRLKQK	985
Dy	1392	-KOKPLP-PLLARVGGNIENVFNTRQRKAFLANVRMGWPQDAFTTQW--LVRLDRGK	1447
Qy	986	TFEINYNIGILFKHIAEIEDENSPDSGDVPGKEGLRIEDVLRIALLIIVQEKXPFVED	1045
Dy	1448	TEKEFKAYVSLFMRHLCPCGADGSETFADGVPREGLSRQOVLTRIGVMSLVKKKVBQBEH	1507
Qy	1046	HPG-----KPVP-----	1053
Dy	1508	INGRWSMPELMPDPADSCKSSRRASSPTKTSTTTPEASATNSPTCKXPATPAFSEKBEGI	1567
Qy	1054	-----SRLERF-----	1060
Dy	1568	RTPLEKEAEHQBPKNRSGIKEMETEADAPSAPSLGERLEPRKIPILEDVPGVUGE	1627
Qy	1061	PGLR-----	1064
Dy	1628	MEPEPYRGDBREKSEDVKGBRELPPGPDRPSNRGREBREETEXPRMFNIADGGFTLHT	1687
Qy	1065	-----SGK---IWKEEHDKIMIRAVLKHGYGRWOAI VDDKELGTQELICKELNFI	1110
Dy	1688	LWQNEERAATISSGKLINEIWHRRHDYLLAGI LVHLHG YARWQDI QNDQAFAI-----INE	1740
Qy	1111	PHISLSAEQAGLOQGSGSNPGAOTNQNPGSVITGNNNASADGAQVNSMFFYRDQMOR	1170
Dy	1741	PF-----KTEANKGNFLEMKN-----	1756
Qy	1171	RLVEFVKRVLLLEKANMYE---YAEBYYGLGSSSIPTPEPEABPKIADTVGVSFIEVD	1227
Dy	1757	----KFLARRFKLLEQALUIEQULRRAAYLNLQEPHPAHAMALHARPABAECIAESHOHLS	1813
Qy	1228	DEMLDGLPKTDPTTSEBGMAAVDNNQARVEIAQHYNQCMCKLLDENARE--SQVAYVNQ	1285
Dy	1814	KESLAG-----NKPANAVLHKVLNQLELLSDMKADVTRLPATLSRI	1855
Qy	1286	PP-STKVNESFRALKS	1300
Dy	1856	PPIAARLQMSERSILS	1871

RESULT 5

```

; RESULT 3
; US-10-104-047-2446
; Sequence 2446, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1el full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:

```

Query Match 26.1%; Score 1874.5; DB 15; Length 979;
Best Local Similarity 43.5%; Pred. No. 1.6e-131;
Matches 423; Conservative 151; Mismatches 258; Indels 141;

[illegible]

RESULT 6
US-10-341-434-61
; Sequence 61, Application US/10341434
; Publication No. US20030215835A1


```

; GENERAL INFORMATION:
; APPLICANT: Origene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 90 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341.434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-341-434-61

Query Match      22.6% Score 1621.5; DB 15; Length 1739;
Best Local Similarity 30.7%; Pred. No. 4.3e-112;
Matches 453; Conservative 229; Mismatches 493; Indels 299; Gaps 43;

Qy 11 RSRKPYNLDSDDDDFVKK-----DRTEQVEAIVRTDAKENACQACG 56
Db 199 RGRKKQSDSDDEDDDEAPKQTRRAAKNVSYKEDDDFE-----TDS-DDLIEMTG 250

Qy 57 ESTNLSNCNTCTYAFHAKCLVPLPKDASVENWRCPEVCVPLNEIDKILDCMPTKSSQ 116
Db 251 EGVDEQDQDNSET-----IEKVLDD--SRLGKKGAT 277

Qy 117 GSS-----DAEPKPIFKV-----OYLKWKGLSYLHCSWVPEKFEQ-----RAYKS 157
Db 278 GASTVYVAIEANGDPDGFTEKDEGEIQYLIKKWGSYIHSWSEESIQQQKVGLKK 337

Qy 158 NHRKTRVNNFHR-----QMESFNNSDDFVAIRPEWTTVDRLACREE-----201
Db 338 LENFKKDEDEIKQWLGKVPEDVEYFNCQELASLNKQIIVERVIAVTKSTKLQTD 397

Qy 202 -----DGELEVLYVKELSYDECYWSESDI--STFONEIQRFKQVNSR-----TRR 246
Db 398 FPAHSRKPAPSNEPEYLCKWMLGPLYSECSWEDEALIGKFQNCIDSFHSRNSKTIPTRE 457

Qy 247 SKVDHNRNPRDQFQDHTPEFKG---LLHPYQLEGLNFRFSWSKOTHVILADEMGLG 303
Db 458 CKAL--KQRP-R-FVALKKQAYLGGENLELRDQLEGLNLWLSHCKNSVILADEMGLG 514

Qy 304 KTIQSIALLASFEENLI--PHLVIAPLSTLRNWEREFATWAPQNMVVMYFGTAQARAVI 361
Db 515 KTIQTISFLYLFHQHLYGPFLLVPLVPLSTLTSQRBFEIWAPEINVVVYIGDLSRNTI 574

Qy 362 REHEFYLSKQKKTKKKKSQISSESQKRIKPDVLITSYEMINLDSAVLKPWKCMIV 421
Db 575 REYEWI-----HSQTKRLKFNALITTYEILLKOKTVLGSINWAPLGV 616

Qy 422 DEGRHLKNKSKLFPSSLTQSSNHRILLTGTPLQNNLDELFLMLHFDLADGKFGSLEBFOE 481
Db 617 DEARLKNDSLLYKTLIDPKSNHRLITGTPLQNSLKLWLSLLHFMPEKFEWDFEE 676

Qy 482 EFKDINOFEQISRLHKMLAPHLRRVKVQVNMKMPKKEILRVLDLSSLOKEYYKAIATR 541
Db 677 DHKG-GRENGYQSLHKVLEPFLLRVRVKVDEKSLPAKVEQILRVEMSAQKQYKWLTR 735

Qy 542 NYQVLT-K-KGAQOISLNNIMELRKVCCHPYMLEGVEP-VIHDANEAFKOLLSCGKLQ 599
Db 736 NYKALAKGTGSTGSLNIVNELKCCNHCYLIKPPENRENGQEILLSLIRSSGKLIL 795

Qy 600 LDKMVKLKEGHRVLYTQFQHMLDLEDYCTHKWQYERIDGKVGGAERQIRIDRFNA 659
Db 796 LDKLLTLRERGNVLIIFSQWVRLDILAEVLTIKHYFPQRLDGSIKGEIRKQALDFNA 855

Qy 660 KNSNKFCLLSTRAGGLGINLATDVIIYDSDNPHADIQAMARHLRGOTNKVMYRL 719
Db 856 DGSDFCFLLSTRAGGLGINLASADTVIIFSDWNPNQDLQAQARHRIQKQVNIYRL 915

```

RESULT 7

```

US-08-973-363-17
; Sequence 17, Application US/08973363
; Publication No. US20030191297A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: AVIAN GHD GENES AND THEIR USE IN METHODS FOR
; TITLE OF INVENTION: SEX IDENTIFICATION IN BIRDS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WENDETH, LIND & PONACK, L.L.P.
; STREET: 2033 K. Street, N.W., Suite 800,
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,363

```

```

; FILING DATE:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/GB96/01341
; FILING DATE: 05-JUN-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: GB 9511439.3
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER: 263/PPNTR1172US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)-721-8200
; TELEFAX: (202)-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1795 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-973-363-17

Query Match 21.98; Score 1570.5; DB 8; Length 1795;
Best Local Similarity 29.54; Pred. No. 3.1e-108; Indels 323; Gaps 47;
Matches 438; Conservative 259; Mismatches 466;

QY 14 RKPVNLDSDDDDFVPKDRTPQEVAVIRTDAKE-----NACQACGSGTSLVSC 64
DB 261 KRQIDSEDEDEDY--DNDKRSRQATVNVSYKEDEEMKTDSDLLLEVCGED----- 312
QY 65 NTCTYAFHAKLPPLPKDASVENWRCEVSPINDEIKILDCEMRPTKSEQGS----- 118
DB 313 -----VQPEDBEFET-----IERVMDC--RVGRKGATGATTIYA 346
QY 119 --SDAEPKPIFVK-----QYLKWKGLSYLHCSWVPEKEFOKAYKSHRLKTRVNNPH 169
DB 347 VEADGDPNAGFERNKEPGDIQYLKWKGLSYLHCSWVPEKEFOKAYKSHRLKTRVNNPH 403
QY 170 RQ-----MESFNSEDDFVAIRPEWTTVDRIACREED---GELEYLVKY 211
DB 404 KQOETKRWLKNASPEDVEYVNCQELTDLHLQYQIVERLIIAHSNQKSAAGLPDYCKW 463
QY 212 KELSDECYWESDIS--TFQNEIQRFQKVN--SRTRSKDVD--HGRNPRDFQOQDHTPEF 268
DB 464 QGLPYSECSWDGALISKFTQCIDVEFSRQSKTTPFKCKVLKQRP--FVALKKQPSY 522
QY 269 LKG-----LLHPYQLEGLNFRFSWKOTHVILADEMGLKTIQSIALLASLPENLI--P 322
DB 523 IGGHEGLELDYQLNGLNWLAHSCWCKGNSCILADEMGLKTIQTIISFLNLPFHEQLYGP 582
QY 323 HLVIAPLSTLRNREBPATWAPQNVVMYFGTAQARAIVREHEFYLSKQKKIKKKSGQ 382
DB 583 FLIVVPLSTLTSQREIQTWASQNAVYVYLDGINSRNIWTHW----- 626
QY 383 ISSSEKQRIKFDVLLTSYEMINDSAVLKPIKWECMIVDEGHLKKNKDSKLFSSLTQVS 442
DB 627 --MHPQTKRLKFNILLTTYEILLKDKAFLGSLNWAFIGVDEAHLKNDSDLLYKTLIDFK 684
QY 443 SNHRIITGTPLONNDELPMHFLDAGKFGSLEBFQEFKQDINQEEQISRLHKMLAPH 502
DB 685 SNHRIITGTPLONSLKELMSLLHFIMPEKFFSWEDFEHGHK--GREYGYASLHKELEPF 743
QY 503 LLRRVKDKVMDKPPKKEILIRVDLSLQKEYKAIPTFNYQVLT--KGAQAI--SLNNIM 561
DB 744 LLRRVKDKVKSIPAKVEQILRMESALQYKQYKWLITRYKALSCKSGKSGTSGFLNIM 803
QY 562 ELKRVCCHPYMLEGVE--PVIHDAEAFKQLLESCKGLQLLDKMMVKLKEQGHRLYITQF 620
DB 804 ELKCCNHCYLIKPPDNNFENYQEAQLHLSRSGKLLLDKLLIRLRERGNRVLIQFSQ 863
QY 621 QHMLDLDLEDYCTHKKQYERIDGKVGAEQRIIDRPNKSNKFKFLLSTRAGGLGINL 680

; FILING DATE:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/GB96/01341
; FILING DATE: 05-JUN-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: GB 9511439.3
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER: 263/PPNTR1172US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)-721-8200
; TELEFAX: (202)-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1795 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-973-363-17

864 VRMLDILAEYLKYPFPQRLDGSIKGEIRKQALDHFNAEGSDFCLLSSTRAGGLGINL 923
681 ATADTVIIYDSDWNPHADLQAMARHRLGQTNKVMYRLINRGITIEERMQLTKKKVLE 740
924 ASADTVIIFDSDWNPQDLQAQARHRIQCKQKQVNIYRLVTKGSVEEDILERAKKQVLD 983
741 HLWVGKLTQKQ-----INOBELEDDIIRYSGKELFASDEDDAGSKGIHYD 785
984 HLVIQRMDDTTGKTVLHTGTSTPSSSTPFNKEELSAILFGAEELFKPEGEQEP-----Q 1038
786 DAAIDKLLDRDLVEAEE--VSVDDEENGFLKAFKAVANFEYIDENEAAALEAQRVAAS 842
1039 EMDIDELIKRAETRENEPGLTVGDE----LLSQFKVANFSNMDDED--IELEPER----- 1088
843 KSSAGNSDRASYWEEL-----LKDQFELHQAELNALGKRKRKRKQVLSIEEDDLAG 894
1089 -----NSRNWEEIIPESQRRRIIEEERQKELEIYMLPRMNRCAQ----- 1129
895 LEDVSDGDESIEAESTDGEAAGQVQVOTGRRPYRRKGRONLEPTPLMEGEGRSFVLGPN 954
1130 ---ISFNGSEGRSRRSRRYSGSDSITERRKPKRGRPTIPRE-----NINQFS 1177
955 QSORAIFVOTLMRYGAGNDFWKEFVR---LKQKTFEEINEYGIILFKHTAEIDENS-- 1009
1178 DAIRIRFIKSYKFG--GPLERLDVAVARDAELVDKSETDLRRLGELVHNGCICKALKONSSG 1236
1010 -----PTFSDGVPKGLRIEDVLRALLIILVQEKVQFVDEHPGKVPFESRI 1056
1237 QERAGGLGKVKGPTF-----RISGVQVNAKLVISHEELA-----PLHKSIPSDP 1282
1057 LER-----PPLRSKGKIMKEEHKIMIRAVLKHGGRQWQAIYDDKELGIQELI----- 1104
1283 EERKTVIIPCHTKAAHFDWDCKEDDSNLLVIGIYEGYSWEMIKMDPDLSTQKILPDD 1342
1105 -----CKELNPFH-----ISLSAAEQAGLOGGSGGSGNPGAQTNQ----- 1141
1343 PDKKPAKQLOQTRADYLIKLKNDKLARKEAQLAGAGNSKRRKTRKNKKNKASKIKKEI 1402
1142 -----PGSVITGNNAASADGAQVNSMFFYRDMORRILVEFKKRVLLLEKAMNYEAEYY 1196
1403 KSDSSPQPSKSEKDEDEEDKNVEM-----KSENKESKKIPLDTPVHTATSE-- 1453
1197 GLGSSSIPTPEBPAPKPIADTVGV-----SFIEVDDMDLGLPKTDPTITSEEIMG 1247
1454 -----PVPISB--ESELHQKTFVCKERMVPVKAALKQLDRPEKGLSEREQLE----- 1500
1248 AAVDNNQARVEIAQHYNQ-----MCKLLDENARESQAQY---VN 1283
1501 ---HTRQCLIKIGDHITTECLKEYNPQIKQWRKNLWIFVSKTFEFDARKLHLYKHAIK 1557
1284 NOPPSTKVNESFRALKSINGNINTILSITSD-----QSKSHEDDTK 1324
1558 KRQESQOHD-----QNISNVNTHVIRNPDVERLAKETTNHDDSSR 1598

RESULT 8
US-08-973-363-16
; Sequence 16, Application US/08973363
; Publication No. US20030191297A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: AVIAN GHD GENES AND THEIR USE IN METHODS FOR
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESS: WENDEROOTH, LIND & POWACK, L.L.P.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patientin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
FILING DATE: US/08/973,363

PRIOR APPLICATION DATA:
FILING DATE: US/08/973,363

PRIOR APPLICATION DATA:
FILING DATE: 05-JUN-1996

PRIOR APPLICATION DATA:
FILING DATE: 06-JUN-1995

ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.

REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER: 263/PPNTIR11720S

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)-721-8200

TELEFAX: (202)-721-8250

TELEX:
INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:
LENGTH: 2108 amino acids

TYPE: amino acid

STRANDEDNESS:
TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-973-363-16

Query Match 21.6%; Score 1555; DB 8; Length 2108;

Best Local Similarity 30.6%; Pred. No. 5.9e-107; Indels 320; Gaps 49;
Matches 430; Conservative 242; Mismatches 411;

QY 7 RLIRSRKPV-----YNLDS-----DDDFVPKOR-----TFQVBAIVRTDAKEN 50
DB 245 RIRPKGKSTGQKRRQLDSEBDDDDYDKGRSRRQATVNVYKEAEE-TKTDS-DD 302
QY 51 ACQACGSTNLVSCNTCTYAFHAKCLVPLKADSVENWRCEVCVPLNEIDKILDCMRP 110
DB 303 LLEVCGED-----VPQTEDEFET-----LEKMD--SRI 330
QY 111 TKSEQSS-----DASPPIFVK-----QYLKWKGLSYLHCSWVPEKFOKAY 155
DB 331 GRKATGASTIYAVEADGPNAGFEKSELGEIQYLKKGWSHIHNTWETETLK--- 387
QY 156 KSNHRLKTRVNNFHRQ-----MESFNSEDDFVAIRPEWTTVDRIACRE 200
DB 388 QONVGMNKLNDYKKQDQETKWLKNASPEDVEYVNCQELTDDLHKQYQIVERIIAHSN 447
QY 201 ED---GELEVLVYKELSYDECYWESESDIS-TFQNEIQRFQOVN-SRTRSKDQVD-HKR 254
DB 448 QKSAAGPYDYCKWQGLPYSECSWEDGALLIAKFPQARIDYFSRNSQSKTTPFFKDCVKLQ 507
QY 255 NPDFQFQDHTPEPLKG---LLHPYQLEGLNFRFSWQTHVILADEMGLKXTIQSIA 310
DB 508 RPR-FVALKKQPSYIGHESLELDYQLNGLNLWHLASHCKNSCILLADENGLKTIOTIS 566
QY 311 LLASLFEENLI--PHLVIAPLSTLRNWEREFATWAPQNVVWVFGTAQARAVIREHEFYLL 368
DB 567 FLNLPFHEHQLYGFPFLRLPLSTLTSQREITQWAPQNAVYVLGDTISRNMTIRTHEW-- 624
QY 369 SKQOKKIKKSKSQISSESQKRIKFDVLLTSVEMINLDSAVLKPWKECWIVDEGHLK 428
DB 625 -----MHPQTKRLKFNILTTTBYELLKDKSFLGGLNWFIVGDEAHLK 668
QY 429 NKDSKLFSSLTQYSSNRHLLTGTPQLNLDLFLMLHFLDAGFKGSLSEFPQEFKDIQ 488
DB 669 NDSLLYRTLDLDFKSNHLLITGTPQLNSLKLWLSLHFMPEKFSWEDFEHSGK-GR 727
QY 489 EEQISRLHKLMAHLLRLRVKDVNMKPPKELLRLVDLSLQKEYKAIPTFRYQVLT 548
DB 728 EYGVASLHKELEPPLRLRVKDVVEKSLPAKVEQILRMEMSAQOQYQYKWLITRNYKALS 787

QY 549 -KGAQOISLNNIMMELRKVCCHPYMLEGVEP-VIHDANEAFKOLLESCGKQLLDKMMVK 606
DB 788 GSKGSTSGFLNIMMELKCCNCYILIKPPDDNEFYNKQALQHLIRSSGKLILLDKLLIR 847
QY 607 LKEQGHRLVITYQFQHMLDLLEDYCTHKKWQVERIDKGVGGAERQIRIDRFNAKNKNKFC 666
DB 848 LRERGNRVLIIFSQVRLMDILAELKYRQFPQRLDGSIKGELRKALDHFNAEGSEDFC 907
QY 667 FLLSTRAGGGINLATADTVIYDSWNPHADLOAMARAHRLGOTWMIYRLINLGTIE 726
DB 908 FLLSTRAGGGINLASADTVIYDSWNPQNDLOAQAARAHRIQKQKVNIYRLVTGSGVE 967
QY 727 ERMMLTKKKVLEHLVVGKLTQN-----INQELDDIIRYSGKELFAS 771
DB 968 EDILERAQKAKWLDHLVIOQMDTGTGTVLHTGSAPSSSTPFNKELLSAILKFAELFKE 1027
QY 772 EDDEAGKSGKIHVDDAAIDKLDRDLVEABE---VSVDDDEENGFLKAFKANFEYIDEN 828
DB 1028 PEGEQEP-----QEMDIDELIKRAETHENEPGLSVGDE---LLSQFVANFNMNMD 1078
QY 829 EAAALEAQRVAEAKSSAGNSDRASYEEL-----LKDKEFLHQAELNALGKRKRS 880
DB 1079 D-IELEPER---NSKN-----WEBIIPEEQRRLEEBERQKELEIYMLPRMNC 1124
QY 881 RKQLVSIEDDLAGLEDVSGSDGSEYAEASTDGEAAGQGVQTRRRPYRRKGRDNLEPTPL 940
DB 1125 AKQ-----ISFNGSEGRSRRRYSGSDSDSISERKPKKGRPTIPRE- 1169
QY 941 MECEGRSFRVLGNQSORAIFVQTLMYRGAGNPDWKEFVPR---LKQTFEENEYGIILF 997
DB 1170 -----NIKGPSDAEIRRFIKSYKKFG-GPLERLDATARDAAELVDKSETDLRLGELV 1220
QY 998 LKHIAEIEIENS-----PTFSDGVKPEGLRIEDVYRIALLILVQEKVKF 1042
DB 1221 HNGCVKALXKSSSGTFRAGRLGVKGPVF-----RISGVQNAKLVIHAHEDEL-- 1269
QY 1043 VEDHPCKVPFPSRILRFPGLRSGKI-----WKEEHDKIMIRAVLKHGGRWQ 1090
DB 1270 IPLHKISPDPE---ER---KQYTIPTCHTKAAHFDIDWGEDDSSNLLIGIYGVGSWE 1322
QY 1091 AIVDDKELGIGIELI-----CKELNPPH---ISLSAAEQALQGN--GSGSNPG 1135
DB 1323 MIKMDPDLSTLTKILPDDPKKQALQOTRADYLIKLSRLAKREAQRLCCGAGS--- 1379
QY 1136 AQTNQPGSVITGNNNASADGAGVNSFYRDMQRLVFEVFKVLLLEKAMNVEABEY 1195
DB 1380 -----KRRKTRAKSKAM-----KSIKVKBEI 1401
QY 1196 YGLGSSSIPTPEEPAEKIADTVGVSVFIEVDDMDLGLPKTDPITSEIM----- 1246
DB 1402 --KSDSPILPSEKSD-----EDDDKINDSKPESKDRSKSVSDAPVHITA 1445
QY 1247 -GAVDNNQARVEIAQHYNQMK 1268
DB 1446 SGEPVPIAESEBELDKTFSICK 1468

RESULT 9

US-08-973-363-20

; Sequence 20, Application US/08973363

; Publication No. US20030191297A1

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: AVIAN GHG GENES AND THEIR USE IN METHODS FOR

; TITLE OF INVENTION: SEX IDENTIFICATION IN BIRDS

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: WENDEROOTH, LIND & PONACK, L.L.P.

; STREET: 2033 K. Street, N.W., Suite 800,

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,363
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: WO PCT/GB96/01341
FILING DATE: 05-JUN-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: GB 9511439.3
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Warren M Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER: 263/PPNTR1172US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)-721-8200
TELEFAX: (202)-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1434 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..1434
OTHER INFORMATION: /note= "The sequence beginning at 1
corresponds to 55 and that ending at 1434 corresponds to
1488"
US-08-973-363-20

Query Match 21.6%; Score 1551.5; DB 8; Length 1434;
Best Local Similarity 32.4%; Pred. No. 5.8e-107;
Matches 408; Conservative 22; Mismatches 389; Indels 243; Gaps 40;
7 RLIRSDRKVV-----YNLDS-----DDDFVFKOR-----TPEQVEAIVRTAKEN 50
117 RIRPKSGKSTGQKQRLDSSEEDDEDYDRGSRQATVNVSYKEABE-TKTDSD-DD 174
51 ACQACGESTNLVSCNTCTAFHAKCLVPLKDAVENWRCPECVPSPLNEIDKILDCMRP 110
175 LLEVCGED-----VPQTEDEFET-----LEKFWD--SRI 202
111 TKSSBOGSS-----DAPKPIFVK-----QYLVKWKGLSYLHCSVWPKFQKAY 155
203 GRKGATGASTTIVAVEADGPNAGFEKSKELGIBIQYLKWKGWSHHTWTETETLK--- 259
156 KSHRLKTRVNNFHRQ-----MESFNSEDDFVAIRPWTTVDRILACRE 200
260 QQNVKGMKLDNKKQDQKQRLKNASPEDVEYNCQBELTDLHKQYQIVERIIAHSN 319
201 ED--GELEYLVKYSDECVWESDIS-TFONEIQRFKDVN-SRTRRSKVDV-HKR 254
320 QKSAAGPYDYCKQGLPYSECSWEDGALLAKFQARIDYFSRNSQKTPPFCKCKVLKQ 379
255 NPRDFQFQDHTPFLKG----LLHPYQLEGFLNFRFSWSKQTHVILADEMGLKTIQSLA 310
380 RPR-FVALKKQPSYIGCHSELSLELDYQLNGLNLAHWSCKGNSCILLADEMGLKTIQIS 438
311 LLASLFEENLI--PHLVIAPLSTLRNWEREPATWAPOMNVMTFTQAQARAVTREHFYL 368
439 FLNYPHEHOLYGFLLRVPFLSTLSQREIQTWAPQMNNAVYLGDTISRNMRITHEW-- 496
369 SKQKKIKKKKSGQISSKQKIKPDVLTSTYEMINLDSAVLKPWKECMIWDEGHRLK 428
497 -----MHPQTKRLKFNILLTTTTEILLKDKSFLGGNNAFVIGVDEAHRLK 540

QY 429 NKDSKLFSSLOYSSNHRILLTGTPLQNNMLDELFMHFLDACKFGSLSEFPQEFKDINO 488
DB 541 NDDSLYRTLIDFKSNHRLITGTPLQNSLKELWSLLHFIMPEKFSFSEWEDFEHKG-GR 599
QY 489 EEQISRLHKLAPHLRLRRVKDKMPPKELILRVDLSSLOKEYKATFTNYOVLTK 548
DB 600 EGYASLHKELEPFLRLRRVKDKVEKSLPAKVEQLRMEMSALQKYKWLTRNYKALS 659
QY 549 -KGAQISLNNIMMELKVCCHPYMLEGVEP-VIHDANEAFKOLLESCGKLQLLDKMMVK 606
DB 660 GSKGSTGFLNIMMELKCCNCHYLKPPDNEFYNKQELQHLIRSSGKLILLDKLLIR 719
QY 607 LKEQHRVLIYTFQHMILLEDYCTHKWQVERIDGKVGGAERQIRIDFNKSNKFC 666
DB 720 LREGRNVLIFSQVMRLDILAEYKYRQFPFQRLDGSINKELRQALDHFNAEGSEDFC 779
QY 667 FLLSTRAGGLGINLATADTVIYDSDWNPHADLQAMARHRLGQTNKVMYRLINRTIE 726
DB 780 FLLSTRAGGLGINLASADTVIYDSDWNPQNDLQOARAHRIQCKQVNIYRLVTKGSVE 839
QY 727 ERMQLTKKKMVLHLVVGKLTQN-----INQELDDIIRYSGKELFAS 771
DB 840 EDILERAKKWLHLVIQRMDDTGTVLHTGSTPSSSTPFNKEELSAILKFGAELFKE 899
QY 772 EDEACKSGKIHYDDAAIDKLLDRDLVEABE---VSDDEENGFLKAFKVANFEYIDEN 828
DB 900 PEGEQEP-----QEMDIDELIKRAETRENEPGLTVGDE---LLSQFKVANFSNMDED 950
QY 829 EAAALEAQRVAASKSSAGNSDRASVWEEL-----LKDKFELHQAELNALGKRKR 880
DB 951 D-IELEPER-----NSRNWEEIIPESQRRRIEERQKELEIYMLPRNRC 996
QY 881 RKQLVSIEDDLAGLDVSDGDESYEASTDGEAAGQVQGTGRRPYRRKGRDNLEPTPL 940
DB 997 AKQ-----ISFNGSEGRSRRSYSGSDSDSITERKPKKGRPTTIPRE- 1041
QY 941 MEGEGRSFVLGPNQORAFVOTLMRYGAGNDFWKEFVPR---LKQKTFEENEYGIILF 997
DB 1042 -----NIKGFSDAEIRRFKSYKFG-GPLERLDAVARDAEALVDKSETDLRRLGELV 1092
QY 998 LXHIAEEDENS-----PTFSDGVKPEGLIEDVLVRIALLILVOEKVKF 1042
DB 1093 HNGICALKNDNSQGRAGRLKVKGPFTF-----RISGVQNAKLVIISHEELA- 1142
QY 1043 VEDHPKCPYFPRILER-----PPGLRSGKIWKEEHDKIMIRAVLKHGYSRQWQIVD 1094
DB 1143 ----PLHKSIIPSDPEERKRYVIPCHTKAAHFDIDWCKEDDSNLLVGIYEGYGSWEMIKM 1198
QY 1095 DKELGIGQELI-----CKELNPPH---ISLSAABQAGLQGGQNGSGSNPQAOTNQ 1141
DB 1199 DPDLSTQKILPDDPKPKQAKLOQTRADYLYLKNLKNLKLARKEAORLAGAGNSKRRKTRN 1258

RESULT 10

US-10-437-963-201156
Sequence 201156, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966

[illegible]

```
Db 749 LRRPQRLDGSWRADURKALDHYNAPGTDFAFLSTSTRAGGLNLTADTVIIFDSD 808
Qy 693 WNPADLQAMARAHRLGQTNKVMYRLINRGTIEERMMLTKKMLVLEHLVVGKLTQN- 751
Db 809 WNPQNDLQAMSAHRIGQTKTVNIYRLVTKGSVEEIEVERAKKVLVLDHLVIQRMDDTGK 868
Qy 752 -----INQELDDIIRYSGKELFASDEDEAGSKIHYYDDAIDKLDRDL 797
Db 869 TVLSKNATAGSVFPFKQELSAIKFGAVELFKKEGEE-----QEPEVD--IDRIL 918
Qy 798 VEAEVSVDDDE--BENGLKAFKVFANFEYIDENEAALAEQORVAESKSSAGNSDRASYW 855
Db 919 MGAETREAEVEMKENELLSFKYVAPDAIDEDKIDAADE-----W 960
Qy 856 EELLKK-----FELHQAEELNALGRKRKRKQ-LVSIIEEDDLAGLEDVSSDGESEYAE 909
Db 961 AAIIPEEDNRILBEERMKELAEWNLAPRQKQIPQVVEDD-----DGDDEED 1011
Qy 910 STDGEAAGQVQGRPRYRKGRDNLEPTPLMEGEGRSFVLGFGNSQORAI FVOTLMRYG 969
Db 1012 -----DTGKKK-KKAVGNF-TIPEIKRPIKSPRK-----FSMPLNRL 1048
Qy 970 --AGNFDWKEFVRLKQKTPPEINEYGIILFKHIAEIDENSPTFSDGVPEKGLRIEDVL 1027
Db 1049 EIAQDAELEHSTDEMKLVESLSE-----ACKAADEFDSNEKNGDAGAAESKKD- 1100
Qy 1028 VRIALLVLQOEKVKF-----VEDHPKRPVPFSPRILE-RFPGLRSKGI- 1068
Db 1101 -----IERKPKFHTCDVNLAKQIERSHAELKPLHEILKSEETKTSFPPANAKLQKGW 1152
Qy 1069 ---WKEHDKIMTRAVLKHGYGRWQAIQVDDKELGIOELI----- 1104
Db 1153 DVDWSRPDDSAIILGVWKYGYGSWEAIKMDPTLGLADKIFIKDKTKKPOGKQLQVRVDYL 1212
Qy 1105 CKELNFPHISLAEQAAGLQGGSGSNPQAQTNQNGPSVITGNNASADGAQVNSMFY 1164
Db 1213 LKLMKDKVKTTEKGRKRKADVPVG-----PEKKKHTNNVPQSGEKKKK- 1259
Qy 1165 YRDMQRLVFEVKRVLILL--EKAMNYEYAEYYG--LGSSSIP----- 1205
Db 1260 -EKKEKNSSLKQDALLSIDKSL-----YGGALDSSAKPFLCVKLCMPVHKYM 1310
Qy 1206 ---TESPEA-----EPKIADTVGVSFIEVDMDGLPKTDPITSEIIMGAADVNNQARV 1257
Db 1311 KKLKEAQEAQNQADEAKYLTRLGDSFLENLETLIKKKPKTN----- 1351
Qy 1258 EIAQHYNQ-----MCKLLDENARESVOAYVNNQPPSTKYNESFRALKSINGNINTILSITS 1313
Db 1352 -IRKWNLYMIFLCKFTLREPGEADRY-----EAXDHKE-----RDREKRRNRGERMDHGGTSK 1383
Qy 1314 DQSKSH-----EDDTKPOLNNVEMKDTABETKPLRGVVDLNVVVEGENIAEASG-SV 1365
Db 1384 DKHKNHHHHHHSKSEKPK-----EAXDHKE-----RDREKRRNRGERMDHGGTSK 1434
Qy 1366 DVQMEBAKEEK 1377
Db 1435 DHHREHHKHDKH 1446
```

RESULT 12

```
US-09-849-602-16
; Sequence 16, Application US/09849602
; Publication No. US20030165834A1
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew J.
; APPLICANT: Old, Lloyd J.
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Chen, Yao-Tseng
; TITLE OF INVENTION: Colon Cancer Antigen Panel
; FILE REFERENCE: L0461/7105(JRV)
; CURRENT APPLICATION NUMBER: US/09/849,602
; CURRENT FILING DATE: 2001-05-04
```

```
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 16
; LENGTH: 1967
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-849-602-16
```

Query Match

Best Local Similarity 30.4%; Score 1492.5; DB 10; Length 1967;
Matches 407; Conservative 228; Mismatches 423; Indels 279; Gaps 48;

```
Qy 100 IDKILDCMRPTKSSQSSDAEPKPIFKVQYLVKMGKLSYLHCSVVPKEFOKAYKSNH 159
Db 40 VKIM--SRSVKQKQSGEEVE-----IEEFYKYKNFSYLHCQWASIEDLSK- ---DK 88
Qy 160 RLKTRVNNF---HRQMESFNNGSDDFVAIRPEWTTVDRII---ACREEDGE--LEVIVKY 211
Db 89 RIQOKIKRFAKQGNKFLSEIEDEL--FNPQYVEVDRIIMDFARSTDDRGEPVTHYLVK 146
Qy 212 KLSYDECWSESISTFQNIQRFKQVNSRTRRSKVDHKNRNP-DFQOFDHTPEFL- 269
Db 147 CSLPYEDSTWERRQDID--QAKIEEFKLMREPETERVE--RPPADDWKKSSESREYKN 202
Qy 270 KGLLHPYQLEGLNLFPSKQTHVILADBMGLGKTIQSTALLASLFEENLI-PHLVIAP 328
Db 203 NKKLREYQLEGVNWLLFNWYNNRNCILADEMGLGKTIQSTITFYIYILKGIHGFVLVIAP 262
Qy 329 LSTLRNWEREFATWAPQMVVMYFGTAQARAVIREHEFYLSKDKQKIKKKKSQIISSEK 308
Db 263 LSTIPNWEREFRTWT-ELNVVYHGSQASRRITQIYEMFKDPQGRVIGS----- 312
Qy 389 QKRIKFDVLLTSYEMINLDSAVLPIKWECEMIVDEGRLKNKDSKLFSSITQVSSNHRIL 448
Db 313 ---YKPHAITITFEMILTDCPELRNIPWRCVVIDEAHRLKNRCKLLEGLKMDLEHKVL 369
Qy 449 LTGTPLONNLDLDELFMVLMHFLDAGKFGSLEEFQEFKDINQEEIQRILHKLMLAPHLRRVK 508
Db 370 LTGTPLQNTVEELLSLHLFLESPFSETTFMQEFGDLKTEEQVQKQLAILKPMMLRLK 429
Qy 509 KDVMDMPKPKELILRLVDSLSSIQEYKAIPTNYQVLTAKGG-AOI-SLNNIMMELRV 566
Db 430 EDVERKMLAPEETIIEVELTNIQKYYRAILEKNFTFLSKGGQANVPLNLTMMELRKC 489
Qy 567 CCHPYMLEGVEPI-----HDANBAFKQL---LESCGKLQLLDKMVKLKEQGHVLI 616
Db 490 CNHPYLINGAEKILBEEFKETHNAESPDLQQAQMIQAAGKVLIDKLLPKLKAGGHRVLI 549
Qy 617 YTFQPHMLDLLEDYCHTKKQYERIDGKVGGAERQIRIDRFNAKNSNKKFCFLSTRAGGL 676
Db 550 FSQWVRCLDILEDYILQRRYPYERIDGRVGRNLRQAAIDRFSPKSDRDFVFLLCSTRAGGL 609
Qy 677 GINLTADTVIIVDSQWNPHADLQAMARAHRLGQTNKVMYRLINRGTIEERMMLTKK 736
Db 610 GINLTADTVIIVDSQWNPHADLQAMARAHRLGQTNKVMYRLINRGTIEERMMLTKK 669
Qy 737 MVLEHLVVGKLIK-----TQNIQEEELDDIIRYSGKELFASDEDEAGSKIHYYDDAA 788
Db 670 LGLDKAVLQSMGSRGNATNGVQLSKKEITEDLLRKAGYALMDEDEGSK- ---FCEED 724
Qy 789 IDKLLDRDLVEAEVSVDDDEENGFLKAFKVFANFEYIDENEAALAEQORVAESKSSAGN 848
Db 725 IDQIL---LRRTHITITIESE--GKGSTTFAKASF-----VAGSNRDI-S 762
Qy 849 SDRASYWEELLKDKFELHQAEELNALGKPK-----RSRKO---LVSTEEDDLAGLEDV 898
Db 763 LDDPNFWQKWK-KAEL-----DIDALNGRNVLVIDTPVRKQTRLYSAVKEDSLEMFSDL 817
Qy 899 SSDGSESYEAESTDGEAAGQVQGTGRPRYRKGRDNLEPTPLMEGEGRSFVLGFGNSQ 957
Db 818 ESDSEKPKCAKPRRQDKSQG-----YARSECFRVEKNLLVYGVGRWTDILSHGRYKR 870
Qy 958 -----RAIFVQTLNRY-----GAGNFDWKEFVRLKQKTPPEINEYGI----- 995
```

Db	871	QLTEQDVETTCRTILVCLNHYKGDENIKSFIWDLITPTADGOTRALVNHSGLSAPVPRG	930	Db	218	HLRIEGHDGTPSEDEYE-----FLIKWVNFSLHCTWEPYNNI-----SMIRGSKV	264
Qy	996	-----LFLKHAEBID-----ENSPTESDGVPKGLR--TEDVLVRIALLILVQEKV	1040	Qy	166	NNFHRQW-----ESFNSEDDFVAI-----REPWTVDRIILACR--EEDGELE	206
Db	931	RKGKKVKAQSTQPVQDADWLASCNPDALFOEDSYKXHLKHHCNKVLLRVMLLYLROEV	990	Db	265	DNHKIVILLDRIRDPDPTTTRIEDIEAMDIKERRKENYEYKQVDRIIVAKHLNSDGSVE	324
Qy	1041	-----KFVB--DHPGKVPFSPRIILFRFGLRSKIKWKEBHDKIMIRAVLKHGVGRWOAI	1092	Qy	207	YLKYKELSYDECWESESDISTF-QWEIQRPXD-VNSRTRRSKDVHDHKNRPDFOQFDH	264
Db	991	IGQADKILGASDESADW---IPEPFHAEVPADWWDKADSKLLIGVFKHYEKYNSM	1047	Db	325	YLKVKQLLYDFTWEASSIIEPIATEIQAFQEREESALSPSRGTNYGNSRPKRYKLEQ	384
Qy	1093	VDDKELGIELICKELNFPFHSLSAAEQAG--LQGGQNGSGG-----	1131	Qy	265	TPBFLK-GULHPVQLSGLNPLRFSKQTHVILADEMGLKTIQSTALLA-----SLPEEN	319
Db	1048	RADPALCFLE----RVGMPDAKAIAAEQRTGDMLAGDGDGEFDEDEDEPYKPTRTPFK	1103	Db	385	QPSYITGGELRDPLQTGVNVMAYLWHKNENGILADEMGLKTVQTVAFLSYLAHSRQHG	444
Qy	1132	-----SNFGAQTQNGPSVITGNNASDGAQVNSMFY-----	1164	Qy	320	LHPLHVIAPLSTLRNWEREFATWAPQWVMVYGTAAQARAVIREHEFYLSKDKKIKKK	379
Db	1104	DEIDEFANSPSEDKESMEIHTATKHS--NABLGQLYWPNSTLTTLRLRLITAYQRS	1161	Db	445	--PFLVVVPLSTVPAWOETLALWDMNCISYLGNTTSRQVIRDIYEFYVDTGOK-----	496
Qy	1165	YRDMORLVEFK-----KRVLLLEKAMNYEABE--YYGLGGSSSI	1204	Qy	380	SGQISSESKQRIKFDVLLTTSYEMINLDSAVLKPIKWECMIVDEGHRLKNKDKSLFSLT	439
Db	1162	YKQOMQOEALMTDRRRRPREBVRALAEAREAIISEKKQKWTTRREADFY-----	1213	Db	497	-----IKFNLLLTYYEVLKDRSVLSNIKWQYMAIDEAHLKNSSESLYEALS	544
Qy	1205	PTEEPEAEPKIADTVGVSFIEVDDEMDLGLPKTDPTITSEIMGAADVNNQARVEIIAQHYN	1264	Qy	440	QYSSNHRILLTGTPLONNLDELPMHFLDAGFGSLEEFQEFKQINQEQISRLHOML	499
Db	1214	-----RVWSTFGVIF-----DPVKQO-----PDWNOFRA-----	1237	Db	545	QPKNSNRLITGTPLQNNIRELAALVDFLMPGKFEIREEINLEAPDEEQEAYIRSLQEHL	604
Qy	1265	QMCKLLDENARESVOAY	1281	Qy	500	APHLARVRKDKVMKMPKPKELILRVDLSSLOKEYKKAIFRNVYQVLTG--KGAQOISLN	557
Db	1238	--FARLDKSDSLEKY	1252	Db	605	QPYILRLRKDKVEKSLPFSKSERILRVELSDQWYKNIILTRNRYRLVTQISSGSLILL	664
RESULT 13				Qy	558	NIMMELRKVCCHPYMLEGVE---PVIHD---ANEAFKOLLESCGKLQLLDKMMVKLKEQ	610
US-10-369-493-2310				Db	665	NIVVELKASNHPYLPDGVESMWQKINSQRRDEVKGLJMNKGWVLKLLSLRLRD	724
; Sequence 2310, Application US/10369493				Qy	611	GHRVLIYTOPHMLDLEDVCTHKWQYERIDKVGGAERQIIRDRNAKNSKNKFCLLS	670
; Publication No. US20030233675A1				Db	725	GHRVLIFSQWVRMLDILGDLRGPYHQRLDGTVPAAVVRTSIDHFNAPNSPFDVLLS	784
; GENERAL INFORMATION:				Qy	671	TRAGGLGINLATADTVIYDSDMNPADLOAMARAHRLGOTNKVMYIRLNRGTIERMM	730
; APPLICANT: Cao, Yongwei				Db	785	TRAGGLGINLATADTVIYDSDMNPADLOAMARAHRLGOTNKVMYIRLNRGTIERMM	844
; APPLICANT: Hinkle, Gregory J.				Qy	731	QLTKKXWVLEHLVVG-----KLKTQINQBELDDIIRYKSKELFASDEDEAGSGKI	782
; APPLICANT: Slater, Steven C.				Db	845	ERARRKMLLEYAIIISLGVTDKQKNKDFSAEELSAILKFGASNMPKAENNQ-----	897
; APPLICANT: Goldman, Barry S.				Qy	783	HYDDAIDKLLDRDLVEAEVSVDDDEENG-----FLKAFKVFANFE-----	823
; APPLICANT: Chen, Xianfeng				Db	898	-----KKLEDMNLDEILSHAEDHDTSDNDVGGASMGGEFLKQFEVTDYKADVSWDDII	950
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF				Qy	824	-----YIDENEAAALEAQ-----RVAAESKSSAGNSDRASYWBELL	859
; FILE REFERENCE: 38-10(52052)B				Db	951	PLTEREKFEEDRLREDEEALKQEIUSRRGNRPYPSSAVESPSYSGTSEKSK-KQML	1009
; CURRENT APPLICATION NUMBER: US/10/369,493				Qy	860	KDKFELHQAEBELNALGKRKRKQLVSIER--DDLAGEDEVSSDGDSEYFAESTDGAAG	917
; CURRENT FILING DATE: 2003-02-28				Db	1010	KDEVILLE--KEIRLL---YRAMIRYGSLEHRYNDIVKYADLTITQDAHVIKKIAADLVAS	1064
; PRIOR APPLICATION NUMBER: US 60/360,039				Qy	918	QGVOTGRRPYRRKGRDNLEPTPLMEGEGRSFVLGFGNQSORAI FVQTLMYRGAGNFWKE	977
; PRIOR FILING DATE: 2002-02-21				Db	1065	RKAVSAAEKDLSNDQSN-----NKSSRKALLITFK--GVKNINAET	1103
; NUMBER OF SEQ ID NOS: 47374				Qy	978	FVRLKQKTFEENEYGLFLKHAIEIDENSPTFSDGVPKGLRIEDVLVRIALLILVQ	1037
; SEQ ID NO 2310				Db	1104	LVORL-----NDLDILY-----DAMPTSGYSNFQIPHW-----	1132
; LENGTH: 1407				Qy	1038	EKVKFVBDHPGKPVFPSRILRFPGLRS-----GKIWKEEHDKIMIRAVLKHGVGRWOAI	1093
; TYPE: PRT				Db	1133	-----RSVHGWSQWGPREDSDMLSGICKHGFGAWLEIR	1166
; ORGANISM: Schizosaccharomyces pombe				Qy	1094	DDKELGLOELICKELNFPFHSLSAAEQAGLQGGQNGSGSNPAGTQNGPNSVITGNNAS	1153
; FEATURE:				Db			
; NAME/KEY: unsure				Qy			
; LOCATION: (1)..(1407)				Db			
; OTHER INFORMATION: unsure at all Xaa locations				Qy			
US-10-369-493-2310				Db			
Query Match				Qy			
Beat Local Similarity				Db			
Matches				Qy			
12 SDRKPVNLD-----SDDDF--VPKDRTFEQVEAIVRTDAKENACQAGES				Db			
118 SRHRTIRDLDEABSVTSEESDSSYGGTPKRSRQKKSNTYVQDEIRFSSRSKGVN				Qy			
59 TNLVSCNTCTYAFHAKCLVPLPKDASVENW-----RCPECVSPINE-----IDKILD				Db			
178 YN-----DAYFESFEDEEEMYEYATEVSEPEDETRDAIDVULD				Qy			
106 CEMRPTKSSQGSDDAEPKPIFVKQYLVKWKGLSYLHCSWVPEKEFQKAYKSNHRLKTRV							

Db 1167 DPELKMCKDIFLE-----DTKQTDNSVPKDKENKE 1197
 QY 1154 ADGAQVNSMYYRDMORRLVEFKVRLVLEKAMNVEYAEYVGLGSSSIPTPEAEAP 1213
 Db 1198 ---KKVPS-----AVHLVRGEGYLLSALREH---HQNFQIKSSPAISTN----- 1235
 QY 1214 KIADTVGVSPFIEVDEMDLGLPKTDPTITSEEINGAAVDNNQARVEIAQHYNQCKLLDEN 1273
 Db 1236 -----GKTOP-----KKQTAN 1246
 QY 1274 ARESQAYVNN-----OPPTKYNESPRALKSINGNINT-----ILSITSDQSKSHE 1320
 Db 1247 RQSGRPNVKSAQKIESATRTSPAISSEKPKFSSKDTKIETPSRQSRQATSPVKSEK 1306
 QY 1321 DDTKPDLLNVEMKDTABETKPLR 1343
 Db 1307 DDGNVSLNAEQKARCKELMYPVR 1329

RESULT 14

US-08-973-363-21
 ; Sequence 21, Application US/08973363
 ; Publication No. US20030191297A1
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: AVIAN GHG GENES AND THEIR USE IN METHODS FOR
 ; IDENTIFICATION IN BIRDS
 ; NUMBER OF SEQUENCES: 39
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: WENDEROTH, LIND & PONACK, L.L.P.
 ; STREET: 2033 K. Street, N.W., Suite 800,
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20006
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30 (BPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/973,363
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO PCT/GB96/01341
 ; FILING DATE: 05-JUN-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9511439.3
 ; FILING DATE: 06-JUN-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warren M. Cheek, Jr.
 ; REGISTRATION NUMBER: 33,367
 ; REFERENCE/DOCKET NUMBER: 263/PPNTIR1172US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)-721-8200
 ; TELEFAX: (202)-721-8250
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 21:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1467 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; FEATURE:
 ; NAME/KEY: Peptide
 ; LOCATION: 1..1467
 ; OTHER INFORMATION: /note= "The sequence beginning at 1
 ; corresponds to 2654 and that ending at 1467 corresponds to
 ; OTHER INFORMATION: 4120"
 ; US-08-973-363-21

Query Match 20.3%; Score 1461.5; DB 8; Length 1467;

Best Local Similarity 32.1%; Pred. No. 3.6e-100;
 Matches 410; Conservative 217; Mismatches 403; Indels 249; Gaps 46;
 QY 7 RLRIIRSDRPVNLDDDDDFVPKXDRTFEQVEAIVRTDAKENACQAGESTNLVSCNT 66
 Db 126 RFSNRQNTVNYNIDYDDDDLESEDD--YGSEAL--SEENVHEA-----SAMP 171
 QY 67 CYAFAHAKLV--PPLKASVENWRCPCEVSPNEIDKILDCMRPTKSEQSSDAEPK 124
 Db 172 QPEDFHGIDIVNHRLLKTSLEEGKVLKTPDLN-----NCKEN----- 210
 QY 125 PIFVKYLVKMGSLVLCCHSWPEKEFOKAYKSNHLK--TRVNNFHRQ-----ME 173
 Db 211 ----YFLKWTDESHLNTW-----ETYSIGQVRGLKRLDNTCKQFIIEQQVRLD 259
 QY 174 SFNNSD-----DFVAIR-----PEWTVTDRIILACRE---EDG--ELEYLVKYLKSYDEC 219
 Db 260 PYVTAEDIEIMDERERRLDEPEHPVPERIIDSQRASLEDGTSQQLYLVKWRRLNDEA 319
 QY 220 YWESSDISTFQNE-IORFKD-VNSSTRSKVDVHKRNPDRDFOQDHTPEFLK-GLLHPY 276
 Db 320 TWENATDIVKLAPEQVKHFQNRNSKILPQYSNNTYSQRPFEKLSVQPPFIKGGELRDF 379
 QY 277 QLEGLNLFKFSWSKOTHVILADEMGIGKTIQSTALLASLF--EENLIPLHVIAPLSLRLN 334
 Db 380 QUTGINWAFKSGDNGILADEMGIGKTVQTVAFISWLIFFARRQNGPHIIVPLSTMPA 439
 QY 335 WEREFAWAPOMVVMYFGTAQARAVIREHEFVLSKQDKKKKKKSGQSSSESKQRIKF 394
 Db 440 WLDTFEKWAPDLNCICYMGQKSRDITREVEFYTNPRAK-----GKTKMF 485
 QY 395 DVLLTSYEMINLDSAVLKPIKWECHIVDEGHLKLNKDSKLFSSLTQYSSNHRILLTGTP 454
 Db 486 NVLLTTYEYLKDRALGSIKQFMAVDEAHLKNAESSLYESLSEFKVANRLITGTP 545
 QY 455 QNNDELFMHFLDAGKFGSLEEFQEEFKDINOEEQISRLHKMLAPHLRLRRVKDVMKD 514
 Db 546 QNNIKELAAALVNLMPGRFTIOEDTFENQDEQEYIHDHRRIOFFILRLRKKDVEKS 605
 QY 515 MPPKELILRLVLDLSSLKQYKAIPTNRYQVLT--KKGGAQISLNNIMMELRKVCCHPYM 572
 Db 606 LPSKTERILRLVELSDVQTEYKNIILTKNYSALTAGAKG-HFSLNIMNELKKSANHPYL 664
 QY 573 LEGVEPVI-----HDANEAFKOLLESCGKQLLDKQMKVKLEQGHVRLVYTOQHML 624
 Db 665 FDNABERVLOKFGDGKWTRENVLRGLIMSGKMWLLDQLLTRLKDGHRVLFPSQVMRL 724
 QY 625 DILEDYCTHKKQYERIDGKVGGAERQIRIDFNKSNKFCFLLSTRAGGLINLATAD 684
 Db 725 DILGDYLSIKGINFORLDGTVPQAORRISIDHFNSPDSNDFVLLSTRAGGLINLATAD 784
 QY 685 TVIYDSDWNPHADLQAMARAHRLGQTNKVMYRLINRGTIERRMQLTKKMKVLEHLVV 744
 Db 785 TVVIFDSDWNPQADLQAMARAHRIQGNHVMYVRLVKDTEVEEVLERARKKMLEVAII 844
 QY 745 G-----KLKTONINQBELDDIIRYGSKELPASEDEAGSKGKIHYDDA---AIDKLL 793
 Db 845 SLGVTGDKNYTKKNEPNAGELSAILKFGAGNFTATDNQK-KLEDLNLDDVLNHAEDHVT 903
 QY 794 DRDLVEAEVSVDDDEENGFLKAFKANF-----EVIDENEAAALEAQRVAESKSSA 846
 Db 904 TPDLGESH---LGGEH---FLKQFVTDYKADIDWDIIPEELKIKQ----- 945
 QY 847 GNSDRASYWEELLKDKFELHQAELNALGKRKRRLQV-SIEEDDLAGLVEDSSDGDES 905
 Db 946 -----DEQKRDQEEYVKEQLEMMNRDNLAKKIKNSVNGDGA---ANSDDDD- 991
 QY 906 YEAESTDGAAGQGVQGTGRPRYRRGRDNLEPTPLMEGGRSF--RVLGFNQ----- 955
 Db 992 ---DSTS-----RSSRRARAN--DMSDICESSEVRALYKAILKFGNLKEILDEL 1035
 QY 956 -SQRAIFVQTLNRYGAGNPDWKEFVPLRKQKTFEINEYGIILFLKHIAEIDENSPTF-- 1012

Db 1036 IADGTLVPKSEFYKE---TYDEMMEAACDCVHEEKN-----RKEILEKLEKHATAYRA 1087
Qy 1013 ---SDGVPEGLRIEDVLRIALILLVQEKVKVEDEHPG-KPVFPSPRIILRFPGLASGK- 1067
Db 1088 KLSGSEIKAENQPKDNPRLRLSL--KKREKAVLFNFKGVKSLNABSLLSRVEDLKYLN 1145
Qy 1068 -----IWKEBHDKIMIRAVLKHGYSRQWQAIVDDKELG 1099
Db 1146 LINSNYKDDPLKPSLGNNTPKPVQNWSSNWTKEDEKLLIGVFKYGSWTQIRDDPFLG 1205
Qy 1100 I-OELICKELNFP--HISLSAAOAGLQONGSG--GSNPGAOTNQNPGSV----- 1145
Db 1206 ITDKIFLNEVHNPAKKSASSDTPTPSKKKGITGS-----SKVPGAHLGRVDYL 1260
Qy 1146 ---ITGNNASADGAQVNS 1161
Db 1261 LSLRGLGLNTKSPSADIGS 1279

RESULT 15
US-10-369-493-1775
; Sequence 1775, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1775
; LENGTH: 1468
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1775

Query Match 20.3%; Score 1461.5; DB 15; Length 1468;
Best Local Similarity 32.1%; Pred. No. 3.6e-100;
Matches 410; Conservative 217; Mismatches 403; Indels 249; Gaps 46;
Qy 7 RLRIRDRKPYNLDSDDDDFVPKDRTPQEVAIVRTDAKENACQAGESTNLVSCNT 66
Db 126 RFSNRQNTVNYIDYSDDDLLSEDD--YGSEAL-----SEENVHEA-----SANP 171
Qy 67 CTYAFHAKCLV--PPLKXASVENWRCEVCYSPLEIDKILDCMRPTKSSQSSDAEPK 124
Db 172 QPEDFHGIDIVNRLKTSLEEGKVLKTPDNL-----NCKEN----- 210
Qy 125 PIFVKQVLVWKGLSYLHCSWPEKEFOKAVKSNHLK--TRVNNFHRQ-----ME 173
Db 211 ----YEFLLKWTDESHLNTW-----EYESIGQVRGLKRLDNYCKPIEDQOVRID 259
Qy 174 SFNNSED-----DFVAIR-----PEWTTVDRIACRE---EDG--ELVLYVKYKLSYDEC 219
Db 260 PYVTAEDIEIMDMERERRRLEDEFEFHVPERIIDSORASLEDGTSQQLVYKWRRLNYDEA 319
Qy 220 WYSESIDISTFQNE-IORFKD-VNSRTRSKVDVHKENPRDFOQDHTPFLK-GLLHPY 276
Db 320 TWENATDIVKLAPQVQKHQFNRENSKILPQVSSNYTSQRPRFKLSVQPPFIKGEURDF 379
Qy 277 QLEGLNLRFSWSKOTHVILADENGLKTIQSIALLASLF--EENLPHLVIAPLSTLRN 334
Db 380 QLTGINWAFNLKSGDNGILLADENGLKTVQTVAFISWLFARRQNGPHIIVPLSTWPA 439
Qy 335 WEREFAWAPQMVVYFGTAQARAVITREHFYLSKDOOKIKKKKSGQISSESKQRIKF 394

Search completed: September 14, 2004, 02:12:18
Job time : 98 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 14, 2004, 01:59:21 ; Search time 53 Seconds
(without alignments)
2511.869 Million cell updates/sec

Title: US-10-049-137-2
Perfect score: 7187
Sequence: 1 MSSLVERLRIRSDRKPYNL.....VDVMEAEKEEKQWVVD 1384

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*
1: Piri.*
2: Piri2.*
3: Piri3.*
4: Piri4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7187	100.0	1384	2 T52301	GYMNOS/PICKLE prot
2	6921.5	96.3	1359	2 B84645	hypothetical prote
3	3114	43.3	1067	2 T06312	hypothetical prote
4	2048.5	28.5	1829	2 T34239	hypothetical prote
5	2032	28.3	1787	2 T20160	hypothetical prote
6	1934.5	26.9	1388	2 T17269	hypothetical prote
7	1572.5	21.9	1711	1 A47392	chromodomain-helic
8	1565.5	21.8	1738	2 C84507	hypothetical prote
9	1536.5	21.4	1465	2 T23056	chromodomain helic
10	1520.5	21.2	1883	2 T13944	chromodomain helic
11	1473	20.5	1388	2 T38720	chromodomain helic
12	1461.5	20.3	1468	1 S30818	hypothetical prote
13	1402	19.5	1373	2 T50107	chromodomain helic
14	1360	18.9	1367	2 T43334	chromodomain helic
15	1332.5	18.5	2957	2 T33152	hypothetical prote
16	1139	15.8	976	2 S35457	SNF2 protein homol
17	1139	15.8	1120	2 S67208	hypothetical prote
18	1135.5	15.8	1143	2 S46122	SNF2 protein homol
19	1131	15.7	1027	2 A56533	chromatin remodell
20	1126	15.7	971	2 S44645	hypothetical prote
21	1120	15.6	1359	2 S49883	nuclear protein ST
22	1114	15.5	1422	2 T18404	chromatin remodell
23	1101.5	15.3	769	2 S35458	SNF2 protein homol
24	1095.5	15.2	1703	2 S15047	SNF2 protein - yea
25	1093.5	15.2	1474	2 T20488	hypothetical prote
26	1052	14.6	1199	2 T37561	probable transcrip
27	1038.5	14.4	1638	2 A42091	transcription acti
28	1030.5	14.3	1022	2 S15078	homeotic gene regu
29	1027	14.3	1647	2 S45525	SNF2beta protein -

30 1024 14.2 1586 2 S39580 HBRM protein - hum
31 1020 14.2 1572 2 S45251 SNF2alpha protein
32 1019.5 14.2 1680 2 T41628 probable transcrip
33 1000 13.9 1257 2 T28937 hypothetical prote
34 987.5 13.7 1613 2 S39059 protein BRG1 - hum
35 922 12.8 778 2 S56293 probable membrane
36 912.5 12.7 861 2 B84885 probable transcrip
37 909.5 12.7 1997 2 F71607 DNA helicase II BR
38 908.5 12.6 603 2 JC4666 lymphocyte specif
39 903.5 12.6 1339 2 A84683 probable SNF2 subf
40 879 12.2 1514 2 S70099 hypothetical prote
41 867 12.1 1288 2 T37528 probable snf2 fami
42 859.5 12.0 1245 2 G84897 hypothetical prote
43 859.5 12.0 1489 2 S60416 DNA helicase IGL15
44 817 11.4 1038 2 T23046 hypothetical prote
45 815.5 11.3 1284 2 T40879 probable helicase

ALIGNMENTS

RESULT 1

T52301
GYMNOS/PICKLE protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Apr-2001
C;Accession: T52301
R;Eshed, Y.; Baum, S.F.; Bowman, J.L.
Cell 99, 199-209, 1999
A;Title: Distinct mechanisms promote polarity establishment in carpels of Arabidopsis th
A;Reference number: Z26025; MUID:20004396; PMID:105335738
A;Accession: T52301
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1384 -ESH>
A;Cross-references: EMBL:AF185578; PIDN:AAF07084.1
A;Experimental source: cultivar Colombia

Query Match 100.0%; Score 7187; DB 2; Length 1384;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1384; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MSSLVERLRIRSDRKPYNLDDSDDDDFVPKQRTPEQVEAIVRTDAKENACQAGSTN	60
Db	1	MSSLVERLRIRSDRKPYNLDDSDDDDFVPKQRTPEQVEAIVRTDAKENACQAGSTN	60
Qy	61	LVSCNTCTYAFHAKCLVPLKDSVENWRCPECVSPLEIDKILDCMRPTKSSEQSSD	120
Db	61	LVSCNTCTYAFHAKCLVPLKDSVENWRCPECVSPLEIDKILDCMRPTKSSEQSSD	120
Qy	121	AEKPIFVKQYLKVKGLSYLHCSWVPEKEFQKAYKSNHRLKTRVNNFHRQMESFNNSD	180
Db	121	AEKPIFVKQYLKVKGLSYLHCSWVPEKEFQKAYKSNHRLKTRVNNFHRQMESFNNSD	180
Qy	181	DFVAIRPEWTTVDRIILACREDEGELEVLVYKLSYDECYWESSEDISTQNEIQRFQDV	240
Db	181	DFVAIRPEWTTVDRIILACREDEGELEVLVYKLSYDECYWESSEDISTQNEIQRFQDV	240
Qy	241	NSRTRSKQVDHKNRPDRFOQDHTPEFLKGLHPYQLEGNLFIRFSWSKQTHVILADEM	300
Db	241	NSRTRSKQVDHKNRPDRFOQDHTPEFLKGLHPYQLEGNLFIRFSWSKQTHVILADEM	300
Qy	301	GLGKTIQSIALLASLFEENLI PHLVIAPLSLRNWEREFATWAPQMVVVFYGTAAQARAV	360
Db	301	GLGKTIQSIALLASLFEENLI PHLVIAPLSLRNWEREFATWAPQMVVVFYGTAAQARAV	360
Qy	361	IRHEFYLSKQDKIKKKSGQISSSKQKRIKFDVLLTYSYEMINLDSAVLKPIKWCMI	420
Db	361	IRHEFYLSKQDKIKKKSGQISSSKQKRIKFDVLLTYSYEMINLDSAVLKPIKWCMI	420
Qy	421	VDSGHRLLKNKDSKLFSLTQYSSNHRILLTGTPQLNNLDELFWLHFLDAGKGSLEEFQ	480
Db	421	VDSGHRLLKNKDSKLFSLTQYSSNHRILLTGTPQLNNLDELFWLHFLDAGKGSLEEFQ	480

481 EEPKIDNOEQISRLHKWLAPELLRRVKKVDMKMPKKEILLRVDLSLQKEYKAIPT 540
Db EEPKIDNOEQISRLHKWLAPELLRRVKKVDMKMPKKEILLRVDLSLQKEYKAIPT 540
QY RNYQVLTCKGGAQISLNNIMMELRKVCCHPYMLGVEPVVHDANEAFKQLLESCKLQLL 600
Db RNYQVLTCKGGAQISLNNIMMELRKVCCHPYMLGVEPVVHDANEAFKQLLESCKLQLL 600
QY DKMMVKLEQGHVRVLIYTOFQHMULDLEDYCTHKWQYERIDGKVGGAERQIRIDRFAK 660
Db DKMMVKLEQGHVRVLIYTOFQHMULDLEDYCTHKWQYERIDGKVGGAERQIRIDRFAK 660
QY NSNKFCLSTRAGGLGINLATADTVIIYSDMNPADLQAMARAHRLGQTNKVMIVRLI 720
Db NSNKFCLSTRAGGLGINLATADTVIIYSDMNPADLQAMARAHRLGQTNKVMIVRLI 720
QY NRGTEIERMMQITKKKVVLEHLVVGKLTQNIQNEELDDIIRYGSKELFASEDDGAKSG 780
Db NRGTEIERMMQITKKKVVLEHLVVGKLTQNIQNEELDDIIRYGSKELFASEDDGAKSG 780
QY KIHVDDAAIDKLDRLDVEABEVSVDDEENGFLKAFKVANFEVIDENEAALEAQRVAA 840
Db KIHVDDAAIDKLDRLDVEABEVSVDDEENGFLKAFKVANFEVIDENEAALEAQRVAA 840
QY ESKSAGNSDRASVWEELLKDFELHQAEBELNALGKRKRKQLVSTEEDDLAAGLEVSS 900
Db ESKSAGNSDRASVWEELLKDFELHQAEBELNALGKRKRKQLVSTEEDDLAAGLEVSS 900
QY DGDESYEAESTDGEAAGGVOTGRPYRRKGRDNLEPTPLMEGGRSFRVLGFNQSORAI 960
Db DGDESYEAESTDGEAAGGVOTGRPYRRKGRDNLEPTPLMEGGRSFRVLGFNQSORAI 960
QY FVQTLMYAGNFWKGFVPLRKQKTFEEINEYGILFLKHTAEBIDENSPFSDGVPKEG 1020
Db FVQTLMYAGNFWKGFVPLRKQKTFEEINEYGILFLKHTAEBIDENSPFSDGVPKEG 1020
QY LRIEDVLRIALLILVOEKVKEVDHPKVPFSPRIILERRPGLRSKGIWKEEHDKIMIRA 1080
Db LRIEDVLRIALLILVOEKVKEVDHPKVPFSPRIILERRPGLRSKGIWKEEHDKIMIRA 1080
QY VLKHGGRWQAI VDDKELGIELCKELNFPFHSLSAAEQAGLQGGSGSNFGAQTNQ 1140
Db VLKHGGRWQAI VDDKELGIELCKELNFPFHSLSAAEQAGLQGGSGSNFGAQTNQ 1140
QY NPGSVITGNNASADGAQVNSMFYRDMQRLVBFVKGRVLLLEKAMNEYAEYGLGG 1200
Db NPGSVITGNNASADGAQVNSMFYRDMQRLVBFVKGRVLLLEKAMNEYAEYGLGG 1200
QY SSSITPEPEAPKPIADTVGVSFTEVDDEMIDGLPKTDPITSEEIMGAADVNNQARVEIA 1260
Db SSSITPEPEAPKPIADTVGVSFTEVDDEMIDGLPKTDPITSEEIMGAADVNNQARVEIA 1260
QY QHYNQMKLLDENARESQAQVNNQPPSTKVNESFRALKSINGNITILSITSDQSKSHE 1320
Db QHYNQMKLLDENARESQAQVNNQPPSTKVNESFRALKSINGNITILSITSDQSKSHE 1320
QY DDTKPDNNVMKQTAETKPLRGVVDLNVVGEENIAEASGVVDVQVMEAKBEEKPKN 1380
Db DDTKPDNNVMKQTAETKPLRGVVDLNVVGEENIAEASGVVDVQVMEAKBEEKPKN 1380
QY MVVD 1384
Db MVVD 1384

RESULT 2

B84645
hypothetical protein At2g25170 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84645
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, J.
euss, D.; Niemman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: B84645
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1359 <STO>
A:Cross-references: GB:AE002093; NID:g6623882; PIDN:AAF19223.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g25170
A:Map position: 2

Query Match 96.1%; Score 6921.5; DB 2; Length 1359;

Best Local Similarity 96.6%; Pred. No. 0;

Matches 1348; Conservative 0; Mismatches 0; Indels 47; Gaps 3;

QY 1 MSSILVERLRIRSDRKPVYNLDDSDDDDFVPKQRTFEQVEAIVRTDAKENACQACGESTN 60
Db 1 MSSILVERLRIRSDRKPVYNLDDSDDDDFVPKQRTFEQVEAIVRTDAKENACQACGESTN 60
QY 61 LVSCNTCTYAFHAKCLVPPLKQASVENWRCPECVSPPLNEIDKILDCMRPTKSQESQSSD 120
Db 61 LVSCNTCTYAFHAKCLVPPLKQASVENWRCPECVSPPLNEIDKILDCMRPTKSQESQSSD 120
QY 121 AEPKPIFVKOYLKWKVGLSYLHCSWVPEKEFOKAYKSNHRLKTRVNNFHRQMFNNSED 180
Db 121 AEPKPIFVKOYLKWKVGLSYLHCSWVPEKEFOKAYKSNHRLKTRVNNFHRQMFNNSED 180
QY 181 DFVAIRPEWTTVDRIILACREEDGELEYLVKYKELSYDECYWESESDISTFQNEIQPKDV 240
Db 170 DFVAIRPEWTTVDRIILACREEDGELEYLVKYKELSYDECYWESESDISTFQNEIQPKDV 229
QY 241 NSRTRRSKVDHKNRPDRFQDFHTPEFLK-----GLLHPYQLGSLNLRFSWS 289
Db 230 NSRTRRSKVDHKNRPDRFQDFHTPEFLKDVMIYLPFAIEGLLHPYQLGSLNLRFSWS 289
QY 290 KQTHVILADEMGLGKTIQSIALLASPEENLI PHLVIAPLSTLRNWEREFATWAPQNVV 349
Db 290 KQTHVILADEMGLGKTIQSIALLASPEENLI PHLVIAPLSTLRNWEREFATWAPQNVV 349
QY 350 MYFCTAQARAVIREHEFYLSKQKKIKKXSGQISSSKQKRIKFDVLLTSYEMINLDSA 409
Db 350 MYFCTAQARAVIREHEFYLSKQKKIKKXSGQISSSKQKRIKFDVLLTSYEMINLDSA 409
QY 410 VLKPIKWECHIVDEGHRLLKNDKSLFSSLTQYSSNHRILLTGTPLONNLDELFLMHFLD 469
Db 410 VLKPIKWECHIVDEGHRLLKNDKSLFSSLTQYSSNHRILLTGTPLONNLDELFLMHFLD 469
QY 470 AGKFGSLEEFQERPKDINQEQISRLHKMLAPHLRRVKQKQVMDMPKKEILLRVDLSS 529
Db 470 AGKFGSLEEFQERPKDINQEQISRLHKMLAPHLRRVKQKQVMDMPKKEILLRVDLSS 529
QY 530 LQKEYYKAIIFTRNYQVLTCKGGAQISLNNIMMELRKVCCHPYMLGVEPVVHDANEAPKQ 589
Db 530 LQKEYYKAIIFTRNYQVLTCKGGAQISLNNIMMELRKVCCHPYMLGVEPVVHDANEAPKQ 589
QY 590 LLESCKLQLLDKMMVKLEQGHVRVLIYTOFQHMULDLEDYCTHKWQYERIDGKVGGA 649
Db 590 LLESCKLQLLDKMMVKLEQGHVRVLIYTOFQHMULDLEDYCTHKWQYERIDGKVGGA 649
QY 650 QRIRIDRPNKNSNKFCLSTRAGGLGINLATADTVIIYSDMNPADLQAMARAHRLG 709
Db 650 QRIRIDRPNKNSNKFCLSTRAGGLGINLATADTVIIYSDMNPADLQAMARAHRLG 709
QY 710 QTNKVMIVRLINRGTEIERMMQITKKKVVLEHLVVGKLTQNIQNEELDDIIRYGSKELF 769
Db 710 QTNKVMIVRLINRGTEIERMMQITKKKVVLEHLVVGKLTQNIQNEELDDIIRYGSKELF 769
QY 770 ASEDDEAGSKGIHYDDAAIDKLDRLDVEABEVSVDDEENGFLKAFKVANFEVIDENE 829
Db 770 ASEDDEAGSKGIHYDDAAIDKLDRLDVEABEVSVDDEENGFLKAFKVANFEVIDENE 829

```
Qy 830 AAALAEQRAEAKSSAGNSDRASYSWELLKDKFELHQAELNALGKRKRGRKOLVISEE 889
Db 830 AAALAEQRAEAKSSAGNSDRASYSWELLKDKFELHQAELNALGKRKRGRKOLVISEE 889
Qy 890 DDLAAGLVSSDGSDEAEASTDGEAAGQVQGTGRPPYRRKGRNLPTPLMEGEGRSFR 949
Db 890 DDLAAGLVSSDGSDEAEASTDGEAAGQVQGTGRPPYRRKGRNLPTPLMEGEGRSFR 949
Qy 950 VLGFNQSORAIFVQTLRMRYGAGNFDWKEFVPRLKQKTFEEINEYGIILFLKHIAEEDENS 1009
Db 950 VLGFNQSORAIFVQTLRM-----YGIILFLKHIAEEDENS 984
Qy 1010 PTFSDGVPKGLRIEDVLRIALLILVQEKVKFVEDHPGKVPFSPRILRPPGLRSGKIW 1069
Db 985 PTFSDGVPKGLRIEDVLRIALLILVQEKVKFVEDHPGKVPFSPRILRPPGLRSGKIW 1044
Qy 1070 KEEDHKIMIRAVLKHGGRWQAI VDDKELGIGQELIKELNPPHISLSAAEQAGQGGNGS 1129
Db 1045 KEEDHKIMIRAVLKHGGRWQAI VDDKELGIGQELIKELNPPHISLSAAEQAGQGGNGS 1104
Qy 1130 GGSNPGAQTNQNGSVITGNNNASADGAQVNSMFYRDMQRRLVEFVKRVLLEKAMNY 1189
Db 1105 GGSNPGAQTNQNGSVITGNNNASADGAQVNSMFYRDMQRRLVEFVKRVLLEKAMNY 1164
Qy 1190 EYAEYYGLGSSSIPTPEPEAKPIADTVGVSVIEVDDEMDGLPKTDPITTSBEIMGAA 1249
Db 1165 EYAEYYGLGSSSIPTPEPEAKPIADTVGVSVIEVDDEMDGLPKTDPITTSBEIMGAA 1224
Qy 1250 VDNNOARVETAOHYNQCKLLDENARESQAAYVNNQPPSTKVNSFRALKSINGNINTIL 1309
Db 1225 VDNNOARVETAOHYNQCKLLDENARESQAAYVNNQPPSTKVNSFRALKSINGNINTIL 1284
Qy 1310 SITSDQSKSHEDDKPDLNINVEMKDTAEETKPLRGVVDVNLNVGEEENIAEASGVVDVKM 1369
Db 1285 SITSDQSKSHEDDKPDLNINVEMKDTAEETKPLRGVVDVNLNVGEEENIAEASGVVDVKM 1344
Qy 1370 EAKEEEKPKNMVVD 1384
Db 1345 EAKEEEKPKNMVVD 1359

RESULT 3
T06312
Hypothetical protein FliC18.100 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999
C:Accession: T06312
R:Bevan, M.; Terry, N.; Ardiles, W.; Buyssehaert, C.; Dasseville, R.; De
  ewes, H.W.; Mayer, K.F.X.; Schueller, C.
  submitted to the Protein Sequence Database, April 1999
A:Reference number: Z15589
A:Accession: T06312
A:Molecule type: DNA
A:Residues: 1-1067 <BEV>
A:Cross-references: EMBL:AL049607; GSPDB:GN00062; ATSP:FliC18.100
A:Experimental source: cultivar Columbia; BAC clone FliC18
C:Genetics:
A:Gene: ATSP:FliC18.100
A:Map position: 4
A:Introns: 40/3; 93/2; 151/2; 223/1; 255/1; 300/3; 326/1; 374/3; 407/2; 453/3; 510/2; 53

Query Match 43.3%; Score 3114; DB 2; Length 1067;
Best Local Similarity 49.3%; Pred. No. 4.2e-150;
Matches 673; Conservative 155; Mismatches 205; Indels 332; Gaps 25;
Qy 1 MSSILVERLRIRSDRKPVYNLDDSDDFVPKORTFEQVEAIVRTDAKENACQAGSTN 60
Db 1 MANILQRLRRRTGPKPDYIEDKLDE-----YIREQVEETGGSNQDC----- 42
Qy 61 LVSCNTCTYAFHAKVLPLPKDASVENRCPECVSPLEIDKILDCEWRPTKSEQGSSD 120
Db 43 -----PLGIEKILDRWRPTASNPNSSD 67
```

```
Qy 121 -ASPKPIFVQYLVKWKGLSYLHCSWVPEKFEQAYKSN-H-RUKTRVNNPHRQWESP-- 175
Db 68 NGTPTLVVVKQYLVKWKGLSYLHCSWVPEQFEKAYKSHPHLKJLVTRFNAAMDVFIA 127
Qy 176 NNSDEDDVAIRPEWTTVDRIILACREDEGELEYLVKYLKELSYDECYWSESDISFQNEIQ 235
Db 128 ENGAHEFIALRPEWKTVDRIIACREGDDGEEYLVKYLKELSYRNSYWSESDISFQNEIQ 187
Qy 236 RFKDVNSRTRRSKDVDRKRNPRDFQQFDHTPEFLKGLLHPYQLLEGLNPLRFSWSKQTHVI 295
Db 188 RFKDINSSRRDKYVENERNRERFKQFDLTPEFLTGTGLHTYQLLEGLNPLRFSWSKKTNI 247
Qy 296 LADEMGLGKTIQSIALLASLFEENLIPHLVIAPLSTURNWEREBEPATWAPQNVVMYFGTA 355
Db 248 LADEMGLGKTIQSIAFLASLFEENLSPLHVAPLSTIRNWEREBEPATWAPNVVMYTGDS 307
Qy 356 QAARAVIREHEFYLSKDKKIKKKSGQISSESQKRIKFDVLLTSYEMINLDSAVLKPPIK 415
Db 308 EARDVIVEHEFYFSEGRK----- 325
Qy 416 WECMIVDEGHRLKNKOSKLFSLTYSSNHRILLTGTPLQNNLDELFLMLHFLDAGKFGS 475
Db 326 ----- 325
Qy 476 LEEFQBEFKDINQEEQISRLHKMLAPHLRRVKQDVNMD-MPPKKELILRLVDLSSLOKEY 534
Db 326 -----RLKDKDLKDKVPPKKELILRLVDMSSQOKEV 355
Qy 535 YKAITFTNYOVLTKGGAQISLANNIMELRKVCCHPYMLGSEVPIHDANEAFKQLLESC 594
Db 356 YKAVITNNYOVLTCKRDAKIS--NVLMLKRVQVSHPYLLPDPFEPFEDANEAFKQLLEAS 413
Qy 595 GKQLQLDKMVKLKEQHRVLIYITQFQMLDLLEDYCTHKKQYERIDGKVGGAERQIRI 654
Db 414 GKQLQLDKMVKLKEQHRVLIYITQFQHTLYLLEDYFTFKNNWYERIDGKISGERQVRI 473
Qy 655 DRFNAKNSKFCFLSTRAGGLGINLATADTVIYDSDMNPHADLQAMARHLRGQTNKV 714
Db 474 DRFNAENSRFCFLSTRAGGIGINLATADTVIYDSDMNPHADLQAMARVHRLGQTNKV 533
Qy 715 MIYRLINRGITERRMOLTKKVKVLEHLVVKLKTQNIQBELDDIIRYSGKELFASDDE 774
Db 534 MIYRLIHKGTVBERMMEITKNRMLLEHLVVGK---QHLQODELDDIIRYSGKELFSEND 590
Qy 775 EAGSGKIHYYDDAIDAKLLDRDLVEAEVSDDEENGFLKAFKVANFEYI-DENEAAAL 833
Db 591 EAGSGKIHYYDDAIEQLLDNRNHVDAVEVSLDDEETDFLKNFKVASFEYVDDENEAAAL 650
Qy 834 EAQRVAABSKSAGNSDRASYSWELLKDKFELHQAELNALGKRKRGRKOLVSTEEDDLA 893
Db 651 E-EAQAIENNSSVRNADRTSHWKDLLKDKYEVOQAELSALGKRKRNGKQ-VMYAEDDL 708
Qy 894 GLEDVSSDGE-----SYEAESTDGEAAGQ-GVQTRRPPYRRKGRNLPTPLME 942
Db 709 GLEBISDEDEYCLDDLKVTSDEBEADEPEAARQKRPVTRPYRKRARDNSSEIPLME 768
Qy 943 GEGRSFRVLGFNQSORAI FVQTLRMRYGAGNFDWKEFVPRLKQKTFEEINEYGIILFLKHIA 1002
Db 769 GEGRYLMLVGFNTERDIFLRTFK-----YGIILFLKHIA 803
Qy 1003 EEIDENSPTF-----SDGVPKGLRIEDVLRIALLILVQEKVKFVEDHPGKVPFSP 1054
Db 804 ENPTDNSTNFKVITAMVYADGVPEKGISDELLVSMTFMMLVKEKCFDNLNHPATPVFSN 863
Qy 1055 RILRPPGLRSGKIWEEDHKIMIRAVLKHGGRWQAI VDDKELGIGQELIKELN--FPH 1112
Db 864 YVTSKY-NLNGAFSKEEDRIILIPAVSKHGIRWAI VDEDEBIGFQEVACKOLNIDFPPP 922
Qy 1113 ISLSAAEQAGLQGGNGSGSNPGAQTNQNGSVITGNNNASADGAQVNSMFYRDMQRRL 1172
Db 923 DTKSA-----RKRI 931
```

```
Qy 1173 VEFVKRVLLEKAMNYEAYEYGLGSSSIPTTEPEAEKPIADITVGVSFIEVDDMLD 1232
Dy 932 CDHVGVKRVKVEDAIKYYAEKIL-----AEQAKAETKGTSEVDAEKML- 976
Qy 1233 GLPKTDPITSEEMGAAYDNNQARVEIAQAHYNQCKLLDENARESQAQYNNQPPSTKVN 1292
Dy 977 ---KNDPITSKKNATAVDNKGQVMAQSYDQS---VNEKSGSFOTYLDIQPLNRMPR 1030
Qy 1293 ESPRALKSINGNINILSITSDQSKSHEDDTKPOLNNVEMKDTAE 1337
Dy 1031 ESFKLEPINEIESTRLSVGTD-----HDVEM-DAAD 1061

RESULT 4
T34239
hypothetical protein F26F12.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34239
R:Wilson, R.; Bentley, D.; Gattung, S.
submitted to the EMBL data library, April 1996
A:Description: The sequence of C. elegans cosmid F26F12.
A:Reference number: Z21493
A:Accession: T34239
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1829 <WIL>
A:Cross-references: EMBL:U55373; PIDN:AAC25894.1; GSPDB:GN00023; CESP:F26F12.7
A:Experimental source: strain Bristol N2; clone F26F12
C:Genetics:
A:Gene: CESP:F26F12.7
A:Map position: 5
A:Intron: 110/3; 441/3; 801/2; 1244/3; 1693/2; 1784/1

Query Match 28.5%; Score 2048.5; DB 2; Length 1829;
Best Local Similarity 32.9%; Pred. No. 9.6e-96;
Matches 518; Conservative 253; Mismatches 474; Indels 331; Gaps 44;

Qy 40 EAVRTDAKEN--ACQAGESTNLVSNTCTYAFHAKLVLPKDA-SVENWRCEC--V 94
Dy 306 EYVKEAPAKQNDPCKIKETENLLDCSCVSHAYCIDPPLTEVPKEETWSPCRCTV 365
Qy 95 SPLNEIDKILDCME-----PTKSEOGSD-----ABPKPIFVKOYLKWKGLSYL 141
Dy 366 KPEHKIEKILCWKKEIPIPEPLEACKESDAMDALKPRKWEPREREPFVKWKLISY 425
Qy 142 HGSWVPEKEFOKAYKSNHRLKTRVN-----FHRQMESFNNSDD-----FVAIR 186
Dy 426 QCSWVSEMLLEVHFRMLILLYWRKNDSDAPPEFEESVTSRHSDNDPYKLRRFYQGIK 485
Qy 187 PEWTTVDRIACRE-EDGELEYLVKYKELSYDECYWE-SESDISTFONEIORF----- 237
Dy 486 PEWQIHRINIHQSYAKSQDYLKWKELSYDQATWERDDSDNIANYEBAIIKYWOHRESK 545
Qy 238 ---KQVNSRTR-----SKDVHKNRPRD-----FOQDHTPEFLK---GLLHPY 276
Dy 546 LNEIDPKQVQKQIAHREAKGLPPKDEKKKKKRDIKRYEYQVDPYVTVETGKLPY 605
Qy 277 QLEGLNLFKFSKQTHVILADEMGLGTIQSIALLASLFEENLI--PHLVIAPISTLRN 334
Dy 606 QLEGLNWLHRHCSNGTDAILADEMGLGKTVQSLTFLYSLMKEGCHCKGPFLIAAPLSTIN 665
Qy 335 WEREFAWAPQNVVYFGTAQARAVIREHEFYLSDQKKIKKKKSGQISSESKQRIKF 394
Dy 666 WEREAEQWCPDFVYVTVYGLRDARVLRHEHESFVEGAVRSGPKASRKMTTEN----MKF 721
Qy 395 DVLTSYEMINLDSVLKPIKWEKMLVDGHLKKNKSKLPSLSLTOYSSNHRILLTGTP 454
Dy 722 HVLLTSYETINMDKTLISIEWGALVVDPAHLKKNQSLFFKQNLNEYTHYHVLITGTP 781
Qy 455 QNNLDELFLMHLFDAGKFGSLEEFQEEFKOINQBEQISRLHKMLAPHLRLRVKDKVMKD 514
Dy 782 QNNLEELFHLNLSKERPNQLEAFTAEPNEISKEQIEKLNLLGPHMLRLKADVLGT 841
```

```
Qy 515 MPPKKEILIRVLDSSLSQKYYKAIPTRYOVL-TKGGGAQISLNNIMMELRKVCCHPYML 573
Dy 842 MPSKSELIVRVELSAMQKWKYKILTFNFDALNVNGSGTQMSLNVLMELKCCNHPYLF 901
Qy 574 EGVEPVIHDANEAPKQ-----LLESCGKLQLLDKMMVKLBKQGHVLIYITQFOHML 624
Dy 902 VKAE-----LEAPKEKNMGVEGTALIKNSGKVFLLQKMLRKLKGGHRLVIFSQMTML 955
Qy 625 DLLEDYCTHKWQYERIDGKVGGAERQIRIDRNNAKSNKFCFLLSFRAGLGINLATAD 684
Dy 956 DIMEDLCYEGYRYERIDGIMGMRQDAIDRVNAPGAQOQIFILLSFRAGLGINLATAD 1015
Qy 685 TVIYDSDNPHADLOAMARAHRLGQTNKVMYIYRLINRGTIERRMOLTKKMKVLEHLV 744
Dy 1016 TVIYDSDNPHNDIQAFSAHRLGQHKVMYIRFVTKSVEEKITSVAKKMLNLHLV 1075
Qy 745 ---GKLTQINIQEELDDIIRYSGKELFASDEDEAGKSG-----KIHYDDAIDK 791
Dy 1076 RAGLGKGEKGTMSKTELDVLRWGTLELFSEDLDAEAGEGSEKKGAAQAEIWDAAVDA 1135
Qy 792 LLDRDLVEAEVSVDDDE-----ENGFLKAFKVFANFEYIDENEAAALAEQAORVAESKSSAG 847
Dy 1136 LLDRSNKEETPAGEDGSEKAEWQNEYLSSFKVASYO-TKETEGQEEEEEETEVIKDEK 1184
Qy 848 NSDRASYWEEILLKDFELHQAELNALGKRSRKLQVSTEED---DLAGLEDVSSDGE 904
Dy 1195 EPD-PDWEKLLKHGHYQDREIELOKLGKGVKQINAYASENMGTDWSQKQNTQDDDD 1253
Qy 905 SYEAESTDGEAAGQGVQVQTRRPY---RRKGRDNIPTPLMEGEGRSFRVLGFGNSQRAIF 961
Dy 1254 N---ESYRGSNCGDGLNSDEDDYDEKKRRRDEEKMPPLMAKVNGQVEILGFNPRQKAF 1310
Qy 962 VQTLNRYGAGNFD-----WKEFVPLKQKTFEEINEYGLFLFKHIAEIDENSPTFSDGV 1016
Dy 1311 YGAVNRMGMPQDQSHQSQW--LVDRLNKSKSVFRAYASLPMRHLCEPGADGHDTFNDGV 1368
Qy 1017 PKEGLRTEDVLRIALLILVOEKVKFVEDHPG-----1048
Dy 1369 PREGLNQHVLRIGLISLVRKVKQEFQEQNGQWSMPEIQDEVLAKAANGSAOGSSRSTP 1428
Qy 1049 -----1049
Dy 1429 KPKEEPKEPEKEDATETVNGATSEPATDAESQNAVPDPEMDTDEAKEPEKPIETEK 1488
Qy 1049 ---KPVPFPPSILE-RPPL-----RSCK---IMKEEDKIMIRAVLKHGCGRW 1089
Dy 1489 PRAARPSFKFICDGGPTLHSLWANBEKVARNKGEYEIWYRRHDYLLAGVHVHGYGRF 1548
Qy 1090 QA---IVDD-----KELGI-----QELICKE-----1107
Dy 1549 QANFNDIINDPRFSVLNPEPFKEVGAETGSDIKAKFMORRFKLIQESLIVIEQLRAAHA 1608
Qy 1108 -----LNPHISLSAAEQAGLQCGSGGSGNPGAGTQNPQNGSVITGNNNASA 1154
Dy 1609 NRHLQPDNVGPLAQRFADLENIAESQANIAKESAGNRNANAVLHKTILVQLDELSDMKA 1668
Qy 1155 DGAQVNSFYRYDMQRLVFEKVRVLLLEKAMNVEYAEYVGLGSSSIPTTEPEAPBK 1214
Dy 1669 DVSRLPSTF-----TQLATVTERLNMTERQILSLTKDADAANRSLVLPDGPFTVPI 1722
Qy 1215 1---ADTVGVGFIEVDDEML---DGLPKTDPITSEEMGAAYDNNQARVEIAQHYNQMC 1267
Dy 1723 LRQQMDGLQPKFAALYSKFMSENGERMEDEPVEAESEGVKQEP-----1767
Qy 1268 KLLDENARESQAQYVNNQPP--STKNVESFRALKSINGNINILSITSDQSKSHEDDTKP 1325
Dy 1768 ---DDETQDSABA-----PPVLSAEVN-----SDSDNDVPSTSAASAVSSE--TAA 1808
Qy 1326 DLNNVEMKDTAEETKP 1341
Dy 1809 DAEPASAEQDPTDEP 1824
```

RESULT 5

T20160
hypothetical protein T14G8.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T20160; T24924
R:McMurray, A
submitted to the EMBL Data Library, November 1995
A:Reference number: Z19231
A:Accession: T20160
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1787 <W1>
A:Cross-references: EMBL:Z67881; PIDN:CAA91798.1; GSPDB:GN00028; CESP:T14G8.1
A:Experimental source: clone C52G5
R:Matthews, P.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z19955
A:Accession: T24924
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1787 <W1>
A:Cross-references: EMBL:Z67884; PIDN:CAA91810.1; GSPDB:GN00028; CESP:T14G8.1
A:Experimental source: clone T14G8
C:Genetics:
A:Gene: CESP:T14G8.1
A:Map position: X
A:Introns: 112/3; 453/3; 597/3; 815/2; 1258/3; 1682/2; 1709/3; 1764/1

Query Match 28.3%; Score 2032; DB 2; Length 1787;
Best Local Similarity 33.5%; Pred. No. 6.3e-95;
Matches 511; Conservative 229; Mismatches 438; Indels 346; Gaps 45;
Qy 19 NLDSDDDDF-VPKDKRTFEQVIAVRTDAKANA--COAGESTNLVSCNTCTVAFHAKC 75
Db 295 NMEQPPGDCWSCPEEHGPDVIVEBEPKANDWYCRICKETSNILLCDTCPSYHAYC 354
Qy 76 LVPLPKDASVNRCPCEV--SPINSEIKILDC-----EMRPTKSQSGSD----- 120
Db 355 IDPLTEPEGEWSCPRCIIEPAQRIEKLISRWKESYSEPLECKEAGEASKDDVFLK 414
Qy 121 --AEPKPIFYQYLVKVKGLSYLCSWVPEKFOKAYKSNHRL---KTRVNN---FHQM 172
Db 415 PPRKMEPREREPFYKWKYLAQWCEMLSDTMDVYFALVRYWYRKVDSENPIFEEST 474
Qy 173 ESNNSDD-----FVAIRPEWTTVDRII-----ACREDEGELEYLVKVKELSYDE 218
Db 475 LSRHSHDHPYKLRERFYQYGVKPEWQIHRINHLVSAKQQD-----YLVKVKELSYEH 530
Qy 219 CYWE-SESISTFQNEIQRF-----KDVNSRTRRSK-----DVPDKR 254
Db 531 ATWERDDTDIANYEDAIKYWHRRMLNDEVPNRVQKMIKQREAKGLGPKEDVTSRR 590
Qy 255 NPRD-----FQDFHTPEFLK---GLLHPYQLEGFLNFRFSWKOTHVILADEMGLGTIQ 307
Db 591 KKREKIDILKKYEVQPDFISFTGTLHPYQLEGFLNFRHCSNGTDAILADEMGLGTIVQ 650
Qy 308 SIALLASLFEENLI--PHLVIAPISTLRNWEREFATWAPQNVVYFCTAQAARAVIREHE 365
Db 651 SLTFLYTLMKGHTKGPFLIAPLSTIINWEREAEIWCPCDPFYVTVYVGDRESRVIREHE 710
Qy 366 FYL-----SKQKKIKKKSGQISSESKQRIKFDVLTSTYEMINLDSAVLKPIKWEQMI 421
Db 711 FSVVDGAVRGGPKVSKIKT-----LENLKFVLLTSYECINMDKAILSSIDWAALV 762
Qy 422 DEGRLLKNDKSLFSSLTQYNSNRHLLTGTPLQNNLDELFLMHFLDAGKFGSLEPQE 481
Db 763 DEARHLKNNQSTFFKRLREYNQYRVLLTGTPLQNNLEELFHLNLFAPDRFNQLESFTA 822
Qy 482 EFKDINOSEQISRLHKMLAPHLRVRVKDKVMKMPKPKELIRVDLDSLOKEYYKAIFTR 541
Db 823 EFSEISKEDQIEKLHLLGLPMHLRLKADVLVTGMPSKQELIVRVLSAMQKKYKYNILTR 882

Qy 542 NYQVL-TKKGQAQISLNNIMMELRKVCCHPYM-----LEGVEPVIHANEAFKOLLBSGC 595
Db 883 NFDALNVKNGTQMSLNIIMELKKCCNHPYLFMKACLEA--PKLKNMGVEGSLIKNAG 940
Qy 596 KLQLLDMMVKLEQGHRLVLYTQFQMLDLEDYCTHKKQYERIDGKVGGAERQIRID 655
Db 941 KVVLLQKMLRKLKDGHRVLI FSGMTMLDILEDPCDVEGYKYERIDGSI TQQRQDAID 1000
Qy 656 RFAKNSNKFCLLSTRAGGLGINLATADTVIYSDNPHADLOAMARHLRGOTNKMV 715
Db 1001 RYNAPGAKQFVLLSTRAGGLGINLATADTVIYSDNPHNDIQAFSRAHRLGQKHVM 1060
Qy 716 IYRLNRGTTIEERMQLTKKMLVHLV-----GKLTQINQBELDDIIRYSGKEIPAS 771
Db 1061 IYRFVTKGSVEERITSVAKKKMLLTHLVVRAGLAKDGKMSKTELDVLRWGTTEELPKE 1120
Qy 772 ED-----DEAGSGK-----IHYDDAAIDKLDRLVEAEVSVDEDEE---NGLPK 815
Db 1121 EEPVEGADGEGTSSKKPNEQEIWDDAAVDFLDNR---KEEGQDGEKKEHTNEVLS 1177
Qy 816 AFKVANFEYIDENEAAALEAQRVAESKSGNSDRASYWEELKDKFELHQAELNALG 875
Db 1178 SPKVATY---NTKEADDDADEDETEVIKEGTEQDNPWEKLLKHVYEQDQETELQKLG 1234
Qy 876 KRRSRKQLVSIED-----DLAGLEDVSSDGESEYAEESTDGEAAGQGVOTGRRPYRKGR 932
Db 1235 KGRVRRQVNYASENMGQWSAQNNQQEEDDGESEYSGDNGELLOTDEYEEERREREER 1294
Qy 933 DNLEPTMEGEGRSFVLGNQSORAI FVOTLMRYGAGNFD-----WKEPVPLKOKTF 987
Db 1295 SEKLP-PLAKVNGQIEVLGNFRQKAFNAVNRWMPQDQLTQSSWQ---VRDLRNKSE 1351
Qy 988 EEINEYGLILKHAIAEIDENSPTFSGVKPEGLRIEDVLRIALLILVQSKV---KF- 1042
Db 1352 KVFKAYSSLFMRHLCEPVVNSDSFMDGVPREGLNRQAVLSRIGLSILRKKVQEFKFN 1411
Qy 1043 ----- 1042
Db 1412 GEWSMPETREKMLATAQAASVSNLPGMIKIKEPIDIDETPMVDQSNITKTEELASEVK 1471
Qy 1043 VDEHPGKPVPSRI-----LERFPLGRSK---IWKEBHKIMIRAVLKHG 1085
Db 1472 VESEPKAPRLPYKPNICDGGYTELHSLWINEEKVARGKEYEIVHRRHDFWLLAAVAYG 1531
Qy 1086 YGRWQAIVDKELGIQELNFPHISLSAAEQAGLQGGSGGSGNPGATQNTQNGSV 1145
Db 1532 YGRYQ-----INFQDI---MNDPKFSIV-----NEPPKQTGADPATN 1565
Qy 1146 ITGNNASADGAQVNSFFYRDMQRRLVFPVKRVLLEKAMNEYEAEEYGLGSSSIP 1205
Db 1566 FADYKN-----KFLARRFKLEQSL----- 1585
Qy 1206 TERPEAPKIADTVGVSVFIEVDDMDLGLPKTDITSEEIMGAADVNNQARV---ETAQ 1261
Db 1586 -----VIEEQLRAAHINKQSPDQVQLAQ 1611
Qy 1262 HYNQMKLLDEN---ARESVQAVYNNQPPSTKVNESFRALKSINGINTILSITSDSKS 1318
Db 1612 HFSELEHTADAHVNIARES-----NNG-----NRNANAILHKCLAQDLD 1650
Qy 1319 HEDDTKPDNLNVMKQTAETKPL 1342
Db 1651 LLSDLKTDV--ARLPATISQVRPV 1672

RESULT 6

T17269
hypothetical protein DKFZp434N231.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T17269
R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, September 1999

A:Reference number: Z18723
A:Accession: T17269
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1388 <POU>
A:Cross-references: EMBL:AL117491
A:Experimental source: adult testis; clone DKFp434N231
C:Genetic(s):
A:Note: DKFp434N231.1

Query Match 26.9%; Score 1934.5; DB 2; Length 1388;
Best Local Similarity 37.2%; Pred. No. 3.9e-90;
Matches 479; Conservative 202; Mismatches 359; Indels 247; Gaps 39;
QY 185 IRPWTVDRL-ACREBDEGLVLYKELSYDECYWE-SESDISTFQNEIQRP----- 237
DB 21 IKPEWMIHRLNHSFDKKGVDHVLKWKDLPYDQCTWEIDDIIDIPYYDNLKQAYWGHRE 80
QY 238 -----KDVNSRTERS KDVKRNPDRFQFQDHTPEFLK-----GLLHPVOL 278
DB 81 LMLGEDTLPLKRLKKGKLRD-DQEKPPDTPIDVPTVKFDKQPWYIDSTGGTLHPVOL 139
QY 279 EGLNLFRESWQKTHVLADWGLGKTIQSIALLASLFE--NLPHLVIAPLSLRNWE 336
DB 140 EGLNLFRESWQKTHVLADWGLGKTIQTVFLVSLYKEGSHGPGYLVASPLSTIINWE 199
QY 337 REPATWAPQMVVMYFGTAQAVIREHEFVLSKDKKIKKKKGQISSESKQRIKFDV 396
DB 200 REFEMWAPDFVVTYTGDKESVIRENEF--SPEDNAIRSGK--KVRPMKEVQIKFHV 255
QY 397 LLTSYEMLNDSAVLKPKIKWCMIVDEGRLKNDKSLFSSLYTOYSSNHRILLTGTPQN 456
DB 256 LLTSYELLTIQAILGSLTEWACLVDDEAHRKNNQSKFRVLSNYKIDYKLLLTGTPLQN 315
QY 457 NLDELFLMLHFLDAGKFGSLSEFEEDKDINQEQISRLHRLMPLAPHLRLRVKQKDMKMP 516
DB 316 NLEELFHLNLFITPERFNNLSEGLFEEADISKEDQIKLHLLGPHMLRLKADVFNKMP 375
QY 517 PKKELILRVDLSQKEYKAIPTFNVOYLTKGGA-OISLNNIMELRVKCCHPYM--L 573
DB 376 AKTELIVRVELSQMKYKFKILTRNFALNSKGGGNQVSLNMMDLKKCCNHPFLPV 435
QY 574 EGVE-PVTHDANAFKQLLESCKQLLDKMMVKLKEQGRVLYLYTQFQMLDLEDYCT 632
DB 436 AAVEAPVLPNGSYDGSSLVKSGKMLKQKMLKLRDEGHRVLFPSQWKMLDLEDPLE 495
QY 633 HKKQYERIDGKVGGAERQIRIDRFNAKNSKFCFLSTRAGGLGINLATADTVIYDSD 692
DB 496 YEGYKIERIDGGITGLRQEAIDRFNAPGAQFCFLSTRAGGLGINLATADTVIYDSD 555
QY 693 WNPADLOAMARAHLRGOTNKVMYRLNRTGIBERMMLTKKQVLEHLVVGK---LKT 749
DB 556 WNPNDIOAFSRAHRIQGNKVMYRFVTRASVEERITQVARKMMLTHLVVREPLGSGS 615
QY 750 QNINQEELEDDIIRVSGKELFASE-----DDEAGKSGK----- 781
DB 616 GSKMTQKQELDLKFTGTELFKDDVEGMMSQQRPVTPIDVQSSKGGNLAASAKKKGST 675
QY 782 -----IHYDDAAILKLDRLVBAEESVDDEE---ENGFLKAFKVFANFEYI 825
DB 676 PPGDNKQVEDSSVIHYDDAAILSKLDLN-----QDATDTELQNNNEVLSLSPKVAQYVR 730
QY 826 DENEAAALEAORVAAESKSGNSDRASYWELLKDKFELHQAEELNALGRKRSKQL- 884
DB 731 BEDGVEEVEREIIKQEE-----NVD-PDYWEKLLRHHYEQQEDLARNLGRKGRIRKQVN 784
QY 885 ---VSIEEDDLALEDVSDGDES-----EAESTDGEAGQGVQGRPRYRK---GRDN 934
DB 785 YNDASQEQE---WQDELSDNQSEYISGEDEDEFEERPEQ-QSGRQSRQLKSDRDK 840
QY 935 LEPTPLMEGEGRSFRVLGNOSQRAIFVQTLMYG-----AGNFDWKEFVPRLKQKTFEE 989
DB 935 LEPTPLMEGEGRSFRVLGNOSQRAIFVQTLMYG-----AGNFDWKEFVPRLKQKTFEE 989

DB 841 PLP-PLLARVGGNIEVLGFNARQKAFNALMRMGMPQDAFNHSHW--LVRDLRGKSEKE 897
QY 990 INEYGLFLKHAIEEIDENSPTFTSDGVPKGLRIEDVLRIALLILYQEKVKFVEDHPGK 1049
DB 898 FRAYVSLFMRHLCEPGADGAETFDAGVPRGLSRQHVLTRIGVNSLVKVKQVEFEHVGK 957
QY 1050 PVFPSRILERFPGLRSKGIWKEEHDKIMIRAVLKHGGRWQAIYDDKELGI---QELIC 1105
DB 958 YSTPDLIPGEGPKKPGEV-----ISSDPTNPVPASPAHLLP 994
QY 1106 KELNFPHISLSAABQAGLOQGGSGNPGAQTNQNPQSVITGNNNASADGAQVNSMYY 1165
DB 995 APLGLPD---KMEQAQLGYMDE---KDPGAQKPRQP----- 1023
QY 1166 RDMQRLVFEVKKREVLLEKAMNVEYAEYVYGLGSSSIPTPEEAPKTIADTVGVSFIE 1225
DB 1024 -----LEVQALPALDRVESEDKH-----ESPASKERARE-----E 1054
QY 1226 VDDEMLDGLPKTDPTITSEEIMGAADVNNQARVEIAQHYNQMKLLDENARESVOAYVNN- 1284
DB 1055 RPEETERAPSPQOLPREVL-----PEKEKILDKLELSLIHSRGDSS 1097
QY 1285 --QPSTKVNESFALKSINGNINTILSITSDQSKSHEDDTKPLNNVEMK---DTAET 1339
DB 1098 ELRPDDTKABEKEPIETQONGD-----KEEDDEGKEDKKGKFKMFNIAD-- 1143
QY 1340 KPLRGVVVDLNVV-EGEENIAEASGSV 1365
DB 1144 ---GGFTLHTLWQNEERAASVSGKI 1166

RESULT 7

A47392 Chromodomain-helicase-DNA-binding protein, CHD-1 - mouse

N:Alternate names: KYBP protein
C:Species: Mus musculus (house mouse)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Dec-2002
C:Accession: A47392; S21568

R:Delmas, V.; Stokes, D.G.; Perry, R.P.
Proc. Natl. Acad. Sci. U.S.A. 90, 2414-2418, 1993

A:Title: A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-1
A:Reference number: A47392; MUID:93211972; PMID:8460153
A:Accession: A47392

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1711

A:Experimental source: S194 plasmacytoma cells

A:Note: sequence inconsistent with the nucleotide translation

A:Note: sequence extracted from NCBI backbone (NCBIN:128272, NCBI:128273)

R:Delmas, V.; Perry, R.P.

submitted to the EMBL Data Library, May 1992

A:Description: KYBP, a mammalian protein that contains the SNF2/SWI2 helicase domain also

A:Reference number: S21568

A:Accession: S21568

A:Molecule type: mRNA

A:Residues: 772-1711 <DE2>

A:Cross-references: EMBL:X66028

C:Superfamily: chromodomain helicase CHD1; chromobox homology

C:Keywords: DNA binding

F:293-336/Domain: chromobox homology <CB1>

F:387-427/Domain: chromobox homology <CB2>

Query Match 21.9%; Score 1572.5; DB 1; Length 1711;

Best Local Similarity 30.7%; Pred. No. 1.2e-71;

Matches 427; Conservative 239; Mismatches 410; Indels 313; Gaps 46;

QY 14 RKPYNLDDSDDDDFVPKQRTTFQVEAIVRTDAKE-----NACQACGESNLVSC 64

DB 209 KRQIDSEDEDEDEY--DNDKSRSRQATVNVSYKEDEMKTDSDLLLEVCGED----- 260

QY 65 NCTCYAFHAKCLVPLPKDASVENWRCPCEVPLNEIKILDCEMRPTKSSQGS----- 118

DB 261 -----VPQPEDEFET-----IERVMDC--RVGRKGATGATTIYA 294

Db 1069 ASIAAFEE-----PSDSVSKQDQNAAGEEDSDKDDDDIIPEGFRKAIIDQERAKEMEDL 1123
Qy 875 -----GKRKRSKQLVSIIEEDDLAGLEDVSSGDSEYAESETDGAAGGV 920
Db 1124 YLPRRKTAQNQNGKRGAGKGKQOADD-----SGDSDYELGS-----DGS 1168
Qy 921 QTGRPRYRRKGRDNLPTPLMEGEGRSPRVLFNGSQORAI FVQTLRMRYGAGNFDWKEFV- 979
Db 1169 GDDGRP-RKGRPTMKE-----KITGFTDAELRRIRSYKKFAPLHRMEALAC 1216
Qy 980 -PRLKQKTFEBINYG-----ILFLKHIAEIDENSPTSFDGVPKE-----GLRIE 1024
Db 1217 DAELQEKPLAELKGLGEMLHDCVQFLHEKKEESKTAATDETCAKORRARATFSVKLG 1276
Qy 1025 DVLVRIALLILVQEKVFEVHPGKVPFPPSILERSFCLRSKGI-----WKEEHDKI 1076
Db 1277 GVSFNAKKLACEQLQ-----PLNEIMPSPPEERQWQSFNIKTRAPVDFVDWGEIBDTK 1331
Qy 1077 MIRAVLKHGYGRWQAI VDDKELGIGIQLICKELNPFHISLSAAEQAGLQGGQSGSGNPGA 1136
Db 1332 LLCGIYQYIGISQWQKLDPTLKLTDKILL----- 1361
Qy 1137 QTNQNGSVITGNNASADGQVNSMFYRDMQRRLVFEVFKRVLLLE----- 1184
Db 1362 -----NDRKPOAKLOTRAEB-----LLKIIKNVELTKGGORRQRPRAS 1403
Qy 1185 KAMNYEABEYVGLGGSSISTEPAPKPIADTVGVSFIEVDDEMLD---GLPKTDPIT 1241
Db 1404 RANDAKAASQ-----SASSTIDAKPHDGEDAAGDARTVA---ESSNSQVDPSNASPHNAPAT 1457
Qy 1242 SEETMGAAVDNQARVETAOHYNQWCKLLDENARESVOAYVYNQPPSTKV-----N 1292
Db 1458 EQH-----GPAKAKKSKARSKTSSASDNNGNKPMHFTANNEPRALEVLGDLDPSTFN 1511
Qy 1293 ESFRALKSINGNINTI-----LSITSQDSKSHEDD 1322
Db 1512 ECKEMRPVKALKALDQDPVSLSDQLOLQTRD 1545

RESULT 11
T38720
chromodomain helicase hrp3 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Dec-2002
C:Accession: T38720
R:Gentles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A:Reference number: Z21797
A:Accession: T38720
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1388 <GEN>
A:Cross-references: EMBL:Z99167; PIDN:CAB16277.1; GSPDB:GN000666; SPDB:SPAC3G6.01
A:Experimental source: strain 972h-; cosmid c3G6
C:Genetics:
A:Gene: SPDB:SPAC3G6.01
A:Map position: 1
A:Introns: 64/3
C:Superfamily: chromodomain helicase CHD1; chromobox homology

Query Match 20.5%; Score 1473; DB 2; Length 1388;
Best Local Similarity 29.3%; Pred. No. 9, 7e-67;
Matches 429; Conservative 215; Mismatches 437; Indels 382; Gaps 42;
Qy 12 SDRKPVYNLD-----SDDDF--VPKQRTFEQVEAIVRTDAKENACOAACES 58
Db 99 SRHRTIRLDDEAESVTSEESDSSYCGTPKRSRQKSNVYQDEIRFSRSRSGVN 158
Qy 59 TNLVSCNTCTAFHAKCLVPLPKASVENW-----RCPECVSPINE-----IDKILD 105
Db 159 YN-----EDAYFSFEEEBEENYEVATEVSEEPEDTRADIVVD 198
Qy 106 CEMRPTKSEQSSDAEPKPIFVKYLVKVLGSLYLHCSWVPEKBFQKAYKSNHRLKTRV 165

Db 199 HRLIEGHGSGTSPSEDYE-----FLIKWVNFSLHCTWEPYNNI-----SMIRSGKV 245
Qy 166 NNFRQM-----ESFNNSEDDFVAI-----RPEWTTVDRIACR-EEDGELE 206
Db 246 DNHIKQVILLDRIREDDPTTTRREDIEAMDIEKREKRENYEYKQVDRIKVAHLNSDGSVE 305
Qy 207 YLVKIKELSDVCWSESDISTF-QNEIQRFKD-VNSRTRRSKQVDHKNRPDRFOQFDH 264
Db 306 YLVKQKLLVDFCTWEASSIIEPIATEIOAFQERESALSPSGTNYGNSRPKRYLEQ 365
Qy 265 TPBFLK-GLLHPYQLEGFLRFWSKQTHVILADEMGLKTIQSIALLA-----SLPEEN 319
Db 366 QPSVITGELDRDQLTGNNWMAVLMHKNENGILLADEMGLKTVQTVAFSLYLAHSLRQH 425
Qy 320 LIHLVIAPISTLRNWEREPATWAPQNVVMYGTQAARAVIREHEHEYSKDOKKIKKK 379
Db 426 --PFLVVPLSTPAPWQETTLALWASDMNCISYLGNTTSRQVIRDIYFVDTGQK----- 477
Qy 380 SGOISSBSKQRIKFDVLLTSYEMINLDSAVLKPikWECMIVDEGHRLLKNKSKLFSLT 439
Db 478 -----IKFNLLITYEVLKDRSVLSNIKWQYMAIDEAHLKXSESSLYEALS 525
Qy 440 QYSSNHRILLTGTPQNNLDELPMHFLPDAGKPSLEEFQEFKQINQBEQISRLHML 499
Db 526 QFKNSNELLITGTPQNNIRELAALVDFLMPGKFEIREINLEAPDEEQEAYIRSLQEHL 585
Qy 500 APHLRRVKQVQKMDPPKELILRDLSSLOQEYKAIETRYNVQVLTG--KGAQOISLN 557
Db 586 QPVILRLKDLKPSKSERILRVELSDQWYKNIILTRNRYRVLQTSISSGSQISLL 645
Qy 558 NIMMELRKVCCHPYMLGVE-----PVTHD---ANEAPKOLLESCGKQLQDDMMVKLEQ 610
Db 646 NIVVELKASNHYPILPGVEESWQKINSQGRDEVVLKGLMNSGKMWLLDKLLSLRRD 705
Qy 611 GHRVLIYTFQHMLDLLEDYCTHKQOYERIDGKVGGAERQIRIDRNKAKNSKFCFLS 670
Db 706 GHRVLIQSVQVVRMLDILGDYLSLRGYPHQLDGTVPAAVRRVTSIDHFNAPNSPDFVLLS 765
Qy 671 TRAGGLGINLATADTVIIYDSDWNPADLOAMARAHGLGOTNKVMYIRLNRGTIERMM 730
Db 766 TRAGGLGINLMTADTVIIFSDWNPQADLOAMARAHQRNHNWVTRLLSKDTIEDVL 825
Qy 731 QLTAKKMWLEHLVVG-----KLQTNINQBELDDIIRYSGKELFASDEDEAGSGKI 782
Db 826 ERARRKMILEYAIISLGVTDKQKSNKDKFSAEELSAILKFGASNMFKENNQ----- 878
Qy 783 HYDDAAIDLLDRDLVEAEVSDDEENG-----PLKAFKVANFE----- 823
Db 879 -----KLEDMNLDILEHAEDHDTSDNVGGASMGGEFLKQFVTDYKADYSWDDII 931
Qy 824 -----YIDENEAALAEQ-----RVAAESKSSAGNSDRASYWEELL 859
Db 932 PLTEREKFEDEDRUREEEALKQBIELSSRRGNRPYPSSAVESPSYSGTSEKSK-KQML 990
Qy 860 KDKPELHOAEELNALGKRKSRKOLVSIIE--DDLAGLEDVSSGDSEYAESETDGAAG 917
Db 991 KDEVILLE--KEIRLL---YRAMIRYGSLEHRYNDIVKYADLTQDAHVIKKIAADLVAS 1045
Qy 918 QGVQVGTGRPYRRKGRDNLPTPLMEGEGRSPRVLFNGSQORAI FVQTLRMRYGAGNFDWKE 977
Db 1046 RKAVSAAAEKDLSDQSN-----NKSRRKALLITFK--GVKNINAE 1084
Qy 978 FVPELKQKTFEENEYIGILFKHIAEIDENSPTSFDGVPKEGLRIEDVLVRIALLILVQ 1037
Db 1085 LVQRL-----NDLDILY-----DAMPTSGYSNFIQPMHV----- 1113
Qy 1038 EKVKFVEDHPGKVPFPSRILERPPLRS-----RSVHGSQWGPREDMSLLSGIKHGFGAWLEIR 1147
Db 1114 ----- 1114
Qy 1094 DDKELGTOELICKELNPFHISLSAAEQAGLQGGQSGSGNPGAQTNPGVSIIGNNAS 1153

Db 1148 DDPELKMCKIFLE-----DTKQTDNSVPKCKENKE 1178
Qy 1154 ADGQVNSMFYRDMQRRLLVEFKVRVLLLEKAMNVEYABEYVGLGGSSIPTEEPBAP 1213
Db 1179 ---KKVPS-----AVHLVRGEYLLSALREH---HQNFQIKSSPAISTN----- 1216
Qy 1214 KIADTVGSFIEVDDEMLDGLPKTDPIITSEIMGAAVDNQAARVIAQHNQMKLLDEN 1273
Db 1217 -----GKTQP-----KKQTAN 1227
Qy 1274 ARBSQVAYVNN-----OPPSPKVNESFRALKSINGNINT-----ILSITSDQSKSHE 1320
Db 1228 RROSGKENVKSQAOKIESATRTSPAISERKKPSSDKTKIETPSREQRSQTASPVKSEK 1287
Qy 1321 DDTKPDILANNVEMKDTAETKPLR 1343
Db 1288 DDGNVSLNAEQAKCKELMYPVR 1310

RESULT 12
S30818
hypothetical protein YER164w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Dec-2002
C:Accession: S30818; S50667
R:Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor, submitted to the EMBL Data Library, February 1993
A:Reference number: S30812
A:Accession: S30818
A:Molecule type: DNA
A:Residues: 1-1468 <MUL>
A:Cross-references: GB:U18917; EMBL:L10718; NID:G603377; PIDN:AAB64691.1; PID:G603404
R:Dietrich, F.S.
A:Description: The sequence of S. cerevisiae cosmids 8229, 9115, 9132, 9981, and lambda submitted to the EMBL Data Library, December 1994
A:Reference number: S50667
A:Accession: S50667
A:Molecule type: DNA
A:Residues: 1-1468 <DIE>
A:Cross-references: EMBL:U18917; NID:G603377; PIDN:AAB64691.1; PID:G603404; GSPDB:GN0000
C:Genetics:
A:Gene: SGD:CHD1; MIPS:YER164w
A:Cross-references: SGD:S0000966; MIPS:YER164w
A:Map position: 5R
C:Superfamily: chromodomain helicase CHD1; chromobox homology
C:Keywords: DNA binding; nucleus
F:195-233/Domain: chromobox homology <CB1>
F:285-327/Domain: chromobox homology <CB2>

Query Match 20.3%; Score 1461.5; DB 1; Length 1468;
Best Local Similarity 32.1%; Pred. No. 4e-66;
Matches 410; Conservative 217; Mismatches 403; Indels 249; Gaps 46;

Qy 7 RLRIERSDKPVNLDSDDDDFVPKDRTEQVEAIVRTDAKENACQACGESINLVSCNT 66
Db 126 RFSNRQNTQVNYNDYSDDDLESEDD--YGESEAL----SEENVHEA-----SANP 171
Qy 67 CTYAFHAKCLV--PPLKDAASVNWRCPECVSPLNEIDKILDCMRPTKSSQSGSDAEPK 124
Db 172 QPEDFHGIDIVNHRILKTSLEEGVLEKTVPDLN-----NCKEN----- 210
Qy 125 PIFVQYLVKWKGLSYLHCSWVPEFQKAYKNHRLK--TRVNNFHRQ-----ME 173
Db 211 ----YEFLIKWTDSEHLHTW-----ETVESIGQVRGLKRLDNYCKQFIIEDQQVRLD 259
Qy 174 SFNNSED-----DFVAIR-----PWTTVDRILACRE---EDG--ELEYIKYKELSYDEC 219
Db 260 PYVTAEDIEIMDRERRLDEFEFHFVPERIIDSQSLASLEDGTSQQLYLVKWRRLNDEA 319
Qy 220 WYSESDISTFQNE-IQRFKD-VNSRTRRSKDVHKNRPDRDFOQDHTPEFLK-GLLHPY 276
Db 320 TWENATDIVKLAPQVKHFQNRNSKILPQVSSNYTSQRPFRFKLSVQPPFIKGELRDF 379

RESULT 13

T50107

Chromodomain helicase hrp1 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

Qy 277 QLEGLNFLBSWSKQTHVILADEMGLGKTIQSTALLASLP--EENLIPHVLVIAPLSLRN 334
Db 380 QLTGIINMAFLMSKGDNGILLADEMGLGKTVQTVAFISWLIFARRQNGPHIIVPLSTMPA 439
Qy 335 WEREFTATWAPQMVVMVFGTAQARAVIRBEHEFYLSKDQKIKKKKSGQISSESKQRIKF 394
Db 440 WLDTFEKWAPDLNCICYMGNKQSRDITREVEFYTNPRAK-----GKTKMKF 485
Qy 395 DVLLTSYEMINLDSAVLKPIKWECMIVDEGHRUKNKDSKLFSSLTQYSSNHRILLTGTP 454
Db 486 NVLLTTTYTILKDRABELGSIKQFMVADHAHLKNAESSLYESLNSFKVANRMLITGTP 545
Qy 455 QNNLDELFWLHFLDAGKCSLEEFQEFKINOBEQISRLHKMLAPHLLRRVKKQVMKD 514
Db 546 QNNIKELAALVNFMPGRFTIDQIDFENQDEQEETIHDLHRRIOFFIURRUKKOVKERS 605
Qy 515 MPPKELILRLVDLSSLSQKEYYKAIFTRNTYQVLT--KKGGAQISLNNIMMELRVKCHPYM 572
Db 606 LPSKTERILRVELSDVQTEYYKNILTKNYSALTAGAKGG-HFSLLNIMMELKASNHPLY 664
Qy 573 LEGVEPVI-----HDANEAFKQLLESGLKQLLDKMMVKLKEQHRVLIYVTFQFHM 624
Db 665 FDNAERVLQKFGDGKMTRENVLRGLIMSGKMLDQLLTRLKDGHRVLIFSQVMRL 724
Qy 625 DLLEDYCTHKWQYERIDGKVGGAERQIRIDRENAKNSKFCFLSTRAGGLGINLATAD 684
Db 725 DILGDYLSKINGINFORLDGTVPSAQRISIDHFNSPDSNDFVLLSTRAGGLGINLATAD 784
Qy 685 TVIIYSDWNPHADLQAMARAHRLGOTNKVMIYRLINRGTIERRMQLTKKMWVLEHLV 744
Db 785 TVVIFSDNPNQADLQAMARAHRIQKNHVMVYRLVSKDTVEEVLERARKMILEVAIL 844
Qy 745 G-----KLQTNINQEBELDIIRYSGKLPASEDEAGKSGKIHYYDA---AIDKLL 793
Db 845 SLGVTGDKYTKKNEPAGELSAILKFGAGNMFTATDNQK-KLEDNLDDVLNAEDHVT 903
Qy 794 DRDLVEABEVSVDDEENGFLKAFKANP-----EYIDENEAALAEQARVAESKSSA 846
Db 904 TPDLGSH---LGEE---FLQFQVTDYKADIDWDDIIPPEELKKLQ----- 945
Qy 847 GNSDRASYWBEILLKDFELHQAELNALGKRKRKQLV-SIEEDDLAAGLEDVSSDGDDES 905
Db 946 -----DEQKRKDEYVKEQLEMMNRDNLAKIKNSVNGDGTA-----ANSDD- 991
Qy 906 YEABSTDGEAAGQVQGTGRRPYRRKGRDNLPTPLMEGGRSF--RVLGFNQ----- 955
Db 992 ---DSTS-----RSSRRARAN-DMSIGESEVRALYKAILKFGNLKEILDEL 1035
Qy 956 -SORAIFVQTLRMRYGAGNFDWKEFVPLKQKTFEEINEYGILFLKHIAEIDENSPFF-- 1012
Db 1036 IADGTLFVKSEFKYGE---TYDEMMEAACDCVHEBEKN-----RKEILEKLEKHATAYRA 1087
Qy 1013 ---SDGVPEGLRIEDVLVRIALLILVOEKVFEVDHPG-KPVFPSPRIERFPGLRSGK- 1067
Db 1088 KLSGSIKAENQKPNLRLSL--KKREKAVLFNFKGVKSNAESLSLRVEDLYKLN 1145
Qy 1068 -----TWKEEHDKIMIRAVLKHGQYGRQWAIIVDDKELG 1099
Db 1146 LINSYKDDPLKPSIGNNTPKPVQNWSSNWTKEDEKLLIGVFKYGVGWSVQTIRDDPFLG 1205
Qy 1100 I-OELLCKELNFP--HISLSAEOAGLQONGSG--GSPNGAQTNPQSGV----- 1145
Db 1206 ITDKIFLNEVHNPAKKSASSDSTTPTPSKKKGITGS-----SKVPGAHLGRVDYL 1260
Qy 1146 ---ITGNNAASADGAOVNS 1161
Db 1261 LSFRLGLNTKSPSADIGS 1279

C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Dec-2002
C;Accession: T50107
R;Seeger, K.; Harris, D.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 2000
A;Reference number: Z25038
A;Accession: T50107
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1373 <SEE>
A;Cross-references: EMBL:ALJ36499; PIDN:CAB66168.1; GSPDB:GN00066; SPDB:SPAC1783.05
A;Experimental source: strain 972h(-); cosmid c1783
C;Genetics:
A;Gene: SPDB:SPAC1783.05
A;Map position: 1
C;Superfamily: chromodomain helicase CHD1; chromobox homology

Query Match 19.5%; Score 1402; DB 2; Length 1373;
Best Local Similarity 28.0%; Pred. No. 3.8e-63;
Matches 395; Conservative 273; Mismatches 467; Indels 274; Gaps 51;

Qy 23 SDDDDFPVKDRTFEQVEAIVRTDAKENACACGSETNLVCNTCTVAFHAKCLVPLPKD 82
Db 145 SDDSDAYDKRHFP--SINSASGTEIRTSLSRSSKGSVN-----YNEQEFYDDFD 192

Qy 83 ASVENWRCPECVPLNEIDKILDCMRPTKSSQSSDAEPKPIFVKQYLKVKGLSYLH 142
Db 193 ---EEEEEVEEQVE--EEVEPIIDFVLNHRKAD--AQDDDPKSSV--QYLKQWQEVSHLH 243

Qy 143 CSWVPEKEFQ--KAYKSNHRUKTRVNNPHRQ-----MESFNSEDDFVAIR----- 186
Db 244 NTWEDYSTLSSVRGK-----KVDNYIKQNIYDREIRDPTTTTFEDIBALDIERERK 296

Qy 187 ---PEWTTVDRIILACR--EEDGELEYLVKVKELSYDECYWESESDI--STFQNEIORF-K 238
Db 297 NMLFEYKIVIRIVASNETNEEGKTEYFVKWQLPYDNCWTB--DADVIYSMAPEVYQFLQ 355

Qy 239 DVNSRTRSKVDVHKRNRPDQFQDHTPEFLK--GLLHPYQLEGFLRFSKQTHVILA 297
Db 356 RENSPLYKYGVFYNTRP--PYRKLEKQPSYIKGGEIRDFQLTGNGWMLWHRNENGILA 414

Qy 298 DEMGLGKTIQSIALLASLFE--ENLI PHLVITAPLSTLNWRBEPATWAPQVNVYGTGA 355
Db 415 DEMGLGKTVQTCFLSYLVHSLKHQHPFLIVVPLSTVPWQETLANWTPDLNSICYTGT 474

Qy 356 QARAVIREHFLSKDOKKIKKSGQISSSESKQRIKFDVLLTSYEMINDSVAVLKPIK 415
Db 475 ESRANIRYEFYLSNRSK-----LKNILLITYEVLKDKQELNNIR 517

Qy 416 WECMIVDEGRLKNKDSKLSFSLTQYSSNHRILITGTPLQNNLDELFLMLHFLDAGRFGS 475
Db 518 WQYLAIDEAHLKNSSESLYETLSQFRTANRLITGTPLQNNLKELASLVNLFMPGKFYI 577

Qy 476 LEEQEEBFKDINQEOISRLHKLAPHLLIRVKVDVMPKPKELILRVDLSLOKEYY 535
Db 578 RDELNFDQPAEQERDIRDLQERLQPFILRLKDKVEKSLPSKSERILRVELSDMQTEWY 637

Qy 536 KAIPTFRTYQVLT--KKGGAQISLNNIMMELRKVCCHPYMLGVEPV-----IHDANFAFK 588
Db 638 KNILTKNVALTGTHTDGRGQSLNLNIVVELKCVSNHPYLPFGAAEKMMGKMTREDTLR 697

Qy 589 QLLSECGQLQLDKMMVKLEQGHRLVLYTQFQHMLDLLEDYCTHKKQWYERIDKVGGA 648
Db 698 GIIMNSGKWLLDKLQRLKHDGHRVLIFSQWRMLNLIGEYMSLRGYNVYQLDGTTPAS 757

Qy 649 ERQIRIDFNKNSKFCFLSTRAGGLGINLATDVIIYVDSWNPHADIQAMARHL 708
Db 758 VRRYSIDHFNAPDSPDFVFLSTRAGGLGINLATDVIIYVDSWNPHADIQAMARHL 817

Qy 709 GQTKNVMYLRINRGTIERMWMQLTKQWVLEHLVVG-----KLKTONINOBEELDDI 760
Db 818 CQKHVNVYRFLSKDTVEEDILERRARRKMLLEYALISLGVTEKSNKSKDKYDAQELSAI 877

Qy 761 IRYGSKELFASDEDEAGKSGKIHYDDAAIDKLLDRDLVEAEV---SYVDEEENGFLKAF 817

Db 878 LKFGASNMFATENQK-KLENMLDD--ILSHAEDRD--SSNDVGGASMGEE---FLKQF 930

Qy 818 KVANFEYIDENEAALAEQARVAEAKSSAGNSDRASVWEELLKDKFELHQAELNALGKR 877

Db 931 EVDYDKAEDLN-----WDDIIPEE-EMERIEBERWIAA 963

Qy 878 KRSRKQLVSTIEEDDLGLEDVSDGDESVEABESTDGEAAGQGVQTGRPRYRRGRDNLEP 937

Db 964 QRAK-----EEERERREEREDEDHPS-----RTYKRTTKS----- 996

Qy 938 TPLMEGGRSFRVLGNQSORAIFVQTLRMRYGAGNFDWKEFVPLRLKQKTEETNEYGIL- 996

Db 997 --ITKQORREEMV--REKEIRLLYRAMIKFLGVDERFDITVKEAELOQATDPKRIYLSA 1052

Qy 997 -FLKHIAEIDE-NSPTFSQGVKPEGLRI-----EDVLVRIALLILVQEKVFE 1044

Db 1053 DMVACADEAVERLGADDTKNQPKKAILIEFKGVKNINAEVTILVRKDLTHLRAYKGLD 1112

Qy 1045 DHPGKPV--PPSRILRFPGLRSCKIWKBEHDKIMIRAVLKHGYSRQWQAIIVDDKELGIOE 1102

Db 1113 --PLKQIIGYPIRSVHSM-----NCWGIKEDSMLLAGINKHGCGWQAIKNDPDLGLHD 1165

Qy 1103 LIC-----KELNF-----PHISLSAAEQAGLOGQSGSGNP 1134

Db 1166 KIFLDEAKNDKESRYVPSAVHLVRRGEYLLSVVREHPDLFVVKTDQ-----P 1212

Qy 1135 GAQTNQNPQSVITGNNASADGAQVNSMFYRYDMORELVEFKRVLLLEKAMNYEAE 1194

Db 1213 TKRYKNRKATPKSTROTTLDGSLNTKSSRTKKK-----EE 1251

Qy 1195 YYGIGSSSIPTEBEPAEPIADTVGVSFIEVDDEMLDGLPKTDPTTSEIMGAADVNNQ 1254

Db 1252 ETNRGDETS-----PEG-----TVGED--EVEEPRQAEPPKRALRSN--GKAASNK 1296

Qy 1255 ARVEIAOHYNQMKLLDENARESVOAVNNQPPSTKVNESFRALKSING-NINTILSITS 1313

Db 1297 TTRNSMKTHSAMDTL---TAAALDAELDNM-SNEKAKEVDHVKSENGESVNE--PNTE 1350

Qy 1314 DQSKSHEDDTKPDLLNVMEMKDTAEETKPL 1342

Db 1351 DLSLETBENT-----TVSDISPL 1368

RESULT 14
T43334
chromodomain helicase hrp1 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Dec-2002
C;Accession: T43334
R;Jin, Y.H.; Yoo, E.J.; Jang, Y.K.; Kim, S.H.; Kim, M.J.; Shim, Y.S.; Lee, J.S.; Choi, I.
Mol. Gen. Genet. 257, 319-329, 1998
A;Title: Isolation and characterization of hrp1+, a new member of the SNF2/SNF2 gene fam
A;Reference number: Z22435; MUID:98179088; PMID:9520266
A;Accession: T43334
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1367 <JIN>
A;Cross-references: EMBL:X99021; NID:g1597720; PIDN:CAA67494.1; PID:g1597721
C;Genetics:
A;Note: hrp1+
C;Function:
A;Description: may act as a negative regulator of cellular growth
C;Superfamily: chromodomain helicase CHD1; chromobox homology

Query Match 18.9%; Score 1360; DB 2; Length 1367;
Best Local Similarity 28.3%; Pred. No. 5.1e-61;
Matches 403; Conservative 247; Mismatches 465; Indels 308; Gaps 52;

Qy 23 SDDDDFPVKDRTFEQVEAIVRTDAKENACACGSETNLVCNTCTVAFHAKCLVPLPKD 82
Db 145 SDDSDAYDKRHFP--SINSASGTEIRTSLSRSSKGSVN-----YNEQEFYDDFD 192

QY 83 ASVENWRCPCVSPINIDKIDCEMPTKSSGSSDABPKPIFVKOYLKVKGLSYLH 142
Db 193 ---EEVEVEQVE---EYEPILIDPVLNHRKAD---AODGPKSY---OYLKQVESHSHL 243
QY 143 CSWYPEKEFO--KAYKSNHRLKTRVNNFHRQ-----MESFNSEDDEFVJR----- 186
Db 244 NTWEDYSTLSSVRGK-----KVDNYIKQNIYDREIREDDPTTFEDIEALDIERERK 296
QY 187 ----PEWTVDRIILACR-EDGELEYLVKVKELSYDECYWESSDI---STFQNEIORF-K 238
Db 297 NMLFEYKIVVERIVASETNEGKTEYFVKVQLPYDNCWE-DSDVIYSMAPEVYQFLQ 355
QY 239 DVNSRTRSDVDHKNRPRQFQDHTPEFLK-GLLHPYQLEGFLNLFKFSKQTHVILA 297
Db 356 RENSPLYPYGVFYNTRP-PYRKLEKQPSYKGEIRDFQLTGINMAYLWHRNENGILA 414
QY 298 DEMGLGKTIOIALLASLFE--ENLIPHVIAPLSTLRNWEREFATWAPQVNVVMVFGTA 355
Db 415 DEMGLGKTQVTCVFLSYLHSLKHQHPFLVPLSTVPANQETLANWTPLNSICYTGT 474
QY 356 QARAVIREHEFYLSKQOKKIKKKSGQISSESQKRIKPDVLTSYEMINLDSAVLKPIK 415
Db 475 ESRAIL-----ESMNSRKLKFNILLTTEYILKDKQELNNIR 511
QY 416 WECMIVDEGHLKKNKSKLPSLSLQYSSNHRILLTGTPLQNNLDELFLMHFLDACKFGS 475
Db 512 WQYLAIDEAHLKNSSESLYETUSQFRTANRLITGTPLQNNLKBLSLNFMPGKFI 571
QY 476 LEEFQEBFKDINOBEQISRLHKMLAPHLRRVKDKVMKMPKKELLRLVDLSLOKEYY 535
Db 572 RDELNFQPAEQBERDTRDQERLOPFLRLKDKVEKSLPSKSERILRVELSDMQTEWY 631
QY 536 KAIFTRNYQVLT--KKGGAQISLNNIMELRKVCCHPYMLEGVEPV-----IHDANEAPF 598
Db 632 KNILTKNYRALTGHTDGRGQLSLNIVVELKVSNNHPYLPFGAAEKWGMKMKTRDTLR 691
QY 589 QLESCEKQLQIDKQWVKLXEOGRVLIYTFQHMLDLLEDYCHKKQWYERIDGKVGGA 648
Db 692 GIIMNSGKVLKDLQRLKHDRHVLIFSQVRLMILGEYMSLRGNYQRLDGTIPAS 751
QY 649 ERQIRIDRFNAKNSKFCFLSTRAGGLGINLATADTVIYDSWNPADLQAMARHL 708
Db 752 VRRVSDHFNAPSPDFVLLSTFAGGLGINLATADTVIIFSDWNPADLQAMARHI 811
QY 709 GQTNKVMYRLINRGRTIERRMQLTKKQVLEHLVWG-----KLKQTNQINQEELEDDI 760
Db 812 GQKNHNVYRFLSKDTEEDILERRARKMILEYAILSLGVTEKSKNSKNDKYDAQELSAI 871
QY 761 IRYGSKELFASDEAGSKGIHYDDAAIDKLDRDLVEABEV---SVDDDEENGFLKAF 817
Db 872 LKFGASNMFKATENQK-KLENMNLDD--ILSHAEDRD--SSNDVGGASMGGEE---FLKQF 924
QY 818 KVANFEVID-----ENEAALAEQRAVAESKSGNSAGNSDRASYWEELLKDK 862
Db 925 EVDYKAEDLNWDDIIEEEMERIEEERMLAQRAKEERERR-----EEBEREN 975
QY 863 FELHQA----EELNALGKRKRKQLYSIEEDDLA-----GLEDVSSDG-DESYEABS 910
Db 976 DEHPSTRYKTKTKSITKQQRREEMVREKIRLLRYAMIKFGLVDRFDTIVKEAELOA 1035
QY 911 TDGEAAGQGVQTRRPRYRGRDNLEPTPLMEGBGRSFRVLGFNQSORAIFVQTLMYGA 970
Db 1036 TDPK-----RIYSLSADMVKACDEAVERLGA 1061
QY 971 GNPDKKEFVRLKQKTFEINEYGILFKHIAEIDENSTPFGVPEKGLRTEDVLRI 1030
Db 1062 D--DTKNKQPR-----KAILIEF-----KGV--KNINATVTLRV 1092
QY 1031 ALLILVOEKVFBEDHPKPV--PPSRILRFPGLRSKGIWKEHDKMIRAVLKHGYGR 1088
Db 1093 KOLTHLHAYKGLD--PLKQIIGYPIRSVHSH-----NCSWIKGSDSMLLAGINKHGF 1145
QY 1089 WQAIIVDDKELGIQELIC-----KELNF-----PHISLSAAEQ 1120

Db 1146 WQAIKNDDPDGLGLHDKILFDEAKNDKESRYVPSAVHLVRGEYLLSVVREHPDLFVVKTDQ 1205
QY 1121 AGLQGGGSGSNPGATQTNQNGSVITGNNNASADGAQVNSMFYRDMQRRLVEFVKRY 1180
Db 1206 -----PTKKYRNKAPTAKSSTRQTTLDGSGISNTKSSRTKKKK----- 1243
QY 1181 LLLKAMVYABEYVGLGSSSIPTEEBEAEPIADTVCVSFIEVDDDEMDLGLPKTDPI 1240
Db 1244 -----BEETNRGDETS-----PEG-----TVGED--EVEEPRQAEPPKRAL 1278
QY 1241 TSEEINGAAVNNQARVEIAQHYNOCKLLDENARESVOAYVNNOPSTKVNESFRALKS 1300
Db 1279 RSNM--GKAASNRKTRTNSMKTHSAMDTL---TAAALDDELNM-SNEAKAEVDHVKS 1332
QY 1301 ING-NINTILSITSDQSKSHEDTKPLNNVEMKDTAEETKPL 1342
Db 1333 ENGESVNE--PNTEDLSLETEENT-----TVSDISPL 1362
RESULT 15
T33152
hypothetical protein T04D1.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33152
R:Davidson, S.; Wohldmann, P.
submitted to the EMBL Data Library, May 1998
A:Description: The sequence of C. elegans cosmid T04D1.
A:Reference number: 221292
A:Accession: T33152
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2957 <DAYS>
A:Cross-references: EMBL:AF067617; PIDN:AAC17559.1; GSPDB:GN00019; CESP:T04D1.4
A:Experimental source: strain Bristol N2; clone T04D1
C:Genetics:
A:Gene: CESP:T04D1.4
A:Map position: 1
A:Introns: 122/3; 293/3; 515/3; 1205/2; 1577/1; 2221/3; 2776/1; 2864/3
Query Match 18.5%; Score 1332.5; DB 2; Length 2957;
Best Local Similarity 29.3%; Pred. No. 3.8e-59;
Matches 368; Conservative 219; Mismatches 386; Indels 281; Gaps 43;
QY 11 RSDRKPVY-----NLDDSDDDDFVP-----KKDRTFEQVEAIVRTDA 47
Db 895 KSYRRPAAVVAKSRKKKSVDDSDDDDFYQGRGKKKGGKKKADPTDEAGVEGAKSDD 954
QY 48 KENACQACGESTNLVSCNTCTYAFHAKILVP-PLKDAIVENWRCPECVSPNLIDKILDC 106
Db 955 EDD-----EFLMKIDTPADPNAMV-----VEKILNM 981
QY 107 EM-----RPTKSSSEQSS-----DAEKP 125
Db 982 RMGKVKKVPETEVAEGSENGEESKVKYKPEKNGESVNGESKSKTSETNE 1041
QY 126 I---FVKQYLVKWKGLSYLHSCWVPEKFKQAYKSNHRLKTRVNNFH-RQMESFNSED 181
Db 1042 IEBEEVEQFLIKWKGSYVHCWEKTAELLEIDK---RVEAKIKRKKVKKSSYIEDDED 1098
QY 182 FVAIRPEWTTVDRIIL-ACREEDGELEYLVKVKELSYDECYWESSDI-STFQNEIORFQOV 240
Db 1099 F---NSDFVIVDRVVDLITDDGQEFVLKIKWSLGYEETWPEIEMIPADKVELWRERQV 1155
QY 241 NSRTR-RSKDVHDKRNPDRFQFDHTPEFLKGL-LLHPYQLEGFLNLFKFSKQTHVILAD 298
Db 1156 IDPAKIREK---QRPEPEWKKMSTSKVWKNNGSLREYQFEGVDVLLICYNAQNCILAD 1212
QY 299 EMGLGKTIOIALLASLFEENLI-PHLVIAPLSTLRNWEREFATWAPQVNVVMVFGTAOA 357
Db 1213 EMGLGKTQIITFLSRIYDYGHPFLVVPVPLSTIQNWREFETWT-DNNAIVHGSAYA 1271

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 14, 2004, 01:51:41 ; Search time 31 Seconds
(without alignments)
2324.680 Million cell updates/sec

Title: US-10-049-137-2

Perfect score: 7187

Sequence: 1 MSSLVERLRISDRKPVNL.....VDVKMEBAKEBKPNMVVD 1384

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2091	29.1	1912	1 CHD4 HUMAN	Q14839 homo sapien
2	2083	29.0	1944	1 CHD3 HUMAN	Q12873 homo sapien
3	2045.5	28.5	1982	1 CHDM_DROME	Q91759 drosophila
4	2032	28.3	1787	1 CHD3_CAEEL	Q22516 caenorhabdi
5	1758	24.5	892	1 CHD3_DROME	O16102 drosophila
6	1621.5	22.6	1739	1 CHD2 HUMAN	O14647 homo sapien
7	1572.5	21.9	1711	1 CHD1_MOUSE	P40201 mus musculu
8	1562.5	21.7	1709	1 CHD1_HUMAN	O14646 homo sapien
9	1494	20.8	2004	1 CHDB_HUMAN	Q9nc88 homo sapien
10	1492.5	20.8	1967	1 CHDA_HUMAN	Q9p2d1 homo sapien
11	1473	20.5	1388	1 HRF3_SCHPO	O14139 schizosacch
12	1461.5	20.3	1468	1 CHD1_YEAST	P32657 saccharomyc
13	1422.5	19.8	2713	1 CHD5_HUMAN	Q8td26 homo sapien
14	1402	19.5	1373	1 HRP1_SCHPO	Q9us25 schizosacch
15	1339	15.8	976	1 SN2L_HUMAN	P28370 homo sapien
16	1135.5	15.8	1129	1 YB95_YEAST	P38144 saccharomyc
17	1131	15.7	1027	1 ISW1_DROME	Q24368 drosophila
18	1126.5	15.7	1009	1 SN2L_CAEEL	P41877 caenorhabdi
19	1120	15.6	1359	1 STH1_YEAST	P32597 saccharomyc
20	1095.5	15.2	1703	1 SNF2_YEAST	P22082 saccharomyc
21	1040.5	14.5	1638	1 BRW_DROME	P25439 drosophila
22	1037	14.3	1647	1 SN24_HUMAN	P51532 homo sapien
23	1024	14.2	1586	1 SN22_HUMAN	P51531 homo sapien
24	922	12.8	778	1 YFK8_YEAST	Q43610 saccharomyc
25	879	12.2	1514	1 YD34_YEAST	Q05471 saccharomyc
26	859.5	12.0	1489	1 YG90_YEAST	P53115 saccharomyc
27	795	11.1	1021	1 SRD1_MOUSE	Q04692 mus musculu
28	789	11.0	1026	1 SRD1_HUMAN	Q9h417 homo sapien
29	784.5	10.9	1085	1 RA36_YEAST	P40352 saccharomyc
30	779.5	10.8	1867	1 M071_YEAST	P32333 saccharomyc
31	750.5	10.4	1131	1 YAB9_YEAST	P31380 saccharomyc
32	722	10.0	1493	1 ERC6_HUMAN	Q03468 homo sapien
33	656	9.1	1849	1 TL72_HUMAN	O14981 homo sapien

ALIGNMENTS

RESULT 1

CHD4_HUMAN

ID CHD4_HUMAN STANDARD; PRT; 1912 AA.

AC Q14839;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Chromodomain helicase-DNA-binding protein 4 (CHD-4) (Mi-2 autoantigen

DE 218 kDa protein) (Mi2-beta).

GN CHD4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP MEDLINE=96017437; PubMed=7575689;

RA Seelig H.P., Moosbrugger I., Ehrfeld H., Fink T., Renz M., Genth E.;

RT "The major dermatomyositis specific Mi-2 autoantigen is a presumed

RT helicase involved in transcriptional activation."

RL Arthritis Rheum. 38:1389-1399(1995).

CC -! FUNCTION: Probable transcription regulator.

CC -! SUBUNIT: Central component of the nucleosome remodelling and

CC -! histone deacetylase (NuRD) repressor complex.

CC -! DISEASE: One of the main antigens reacting with anti-Mi-2 positive

CC sera of dermatomyositis.

CC -! SIMILARITY: Belongs to the SNF2/RAD54 helicase family.

CC -! SIMILARITY: Contains 2 chromo domains.

CC -! SIMILARITY: Contains 2 PHD-type zinc fingers.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X86691; CAA60384.1; -.

DR TRANSFAC; T05116; -.

DR Genem; HGNC:1919; CHD4.

DR MIM; 603277; -.

DR GO; GO:0004403; F:ATP dependent DNA helicase activity; TAS.

DR GO; GO:0003677; F:DNA binding; TAS.

DR GO; GO:0008270; F:zinc ion binding; TAS.

DR GO; GO:0007001; P:chromosome organization and biogenesis (sen. .; TAS.

DR GO; GO:0006357; P:regulation of transcription from Pol II pro. .; TAS.

DR InterPro; IPR000953; Chromo.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR002464; DEAD box.

DR InterPro; IPR001650; Helicase_C.

DR InterPro; IPR000330; SNF2_N.

DR Pfam; PF00385; chromo; 1.

DR Pfam; PF00271; helicase_C; 1.

DE	240 kDa protein) (Mi2-alpha).	FT	DOMAIN	631	673	CHROMO 2.
GN	CHD3.	FT	NP_BIND	761	768	ATP (POTENTIAL).
OS	Homo sapiens (Human).	FT	SITE	883	886	DEAH BOX.
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	FT	DOMAIN	206	221	POLY-ALA.
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	FT	DOMAIN	243	246	POLY-PRO.
ON	NCBI_TaxID=9606;	FT	DOMAIN	355	358	POLY-LYS.
OX	[1]	FT	DOMAIN	434	446	POLY-GLU.
RP	SEQUENCE FROM N.A.	FT	DOMAIN	697	703	POLY-LYS.
RC	TISSUE=Fetal;	FT	CONFLICT	121	126	GEDGG -> PHFOQK (IN REF. 2).
RX	MEDLINE=97470991; PubMed=9326634;	FT	CONFLICT	309	312	MISSING (IN REF. 2).
RA	Woodage T., Baerai M.A., Baxevanis A.D., Hieter P., Collins F.S.;	FT	CONFLICT	653	653	W -> G (IN REF. 2).
RT	"Characterization of the CHD family of proteins.";	FT	CONFLICT	1944	1944	AA; 220691 MW; 0712E6F1D5F7D335 CRC64;
RL	Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997).	SQ	SEQUENCE			
RN	[2]					
RP	SEQUENCE OF 121-654 FROM N.A.					
RC	TISSUE=Thymus;					
RX	MEDLINE=96013633; PubMed=7560064;					
RA	Ge Q., Nilasena D.S., O'Brien C.A., Frank M.B., Targoff I.N.;					
RT	"Molecular analysis of a major antigenic region of the 240 kD protein	QY	52	COACGESTNLVSCNTCTVAFHAKLVLPPKDAVENWRCPEVCSP--NEIDKILDCMR	109	
RT	of Mi-2 autoantigen.";	Db	459	CRVCKDGGELCCDACISSYHHCNLPPLDPINGEWLCPRCTCPVLKGRVQKILHWVG	518	
RL	J. Clin. Invest. 96:1730-1737(1995).	QY	110	-----PTKSBEQSSDA-EPKPI---FVKQYLVKMGSLSYLHCSWVPEKQKAYKSNH	159	
CC	-!- FUNCTION: Probable transcription regulator.	Db	519	BPVAVPAPQADGNPDVPPRPLQGRSEREFVVKWGLSYWHCSWAKELQLEIFHLVMY	578	
CC	-!- SUBUNIT: Central component of the nucleosome remodelling and					
CC	histone deacetylase (NuRD) repressive complex.					
CC	-!- SUBCELLULAR LOCATION: Nuclear.	QY	160	RLKTRVNNFHRQME--SFNNSDD-----FVAIRPEWTTVDRI--	196	
CC	-!- DISEASE: One of the main antigens reacting with anti-Mi-2 positive					
CC	sera of dermatomyositis.					
CC	-!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.	Db	579	RNYQRKNDMPPLDYGGSDGSKDRKVKDPHYAEMEEKYKYGKPEWMTVHRIIN	638	
CC	-!- SIMILARITY: Contains 2 PHD-type zinc fingers.	QY	197	ACREEDGEVLVYKVELSYDECYW-ESESDISTFQNEIQ---RFDVNSRTRRSKVDH	252	
CC	-!- SIMILARITY: Contains 2 PHD-type zinc fingers.					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	Db	639	HSVDDKKGNHYLVKWRDLPYDQSTWEEDMMIPEYEEHKQSWWRHRELIMGDDPAQPKY	698	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	QY	253	KENRDFQ-----QFDHTPEFLK---GLLHPYQLEGFLNFRFSWQTHVI	295	
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way	Db	699	KKKKELQDGPSSPTNDPTVKYETQPRFITATGTLHMVQLEGLNMLRFSWAQGTDTI	758	
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	QY	296	LADENGLKTTQSIALLASLFEENLI--PHLVIAPLSTLRNWEREFATWAPQNVVMVFG	353	
CC	or send an email to license@isb-sib.ch).	Db	759	LADENGLKTTQTVIFVLSLYKEGTGKGFPLVSAPLSTIINWEREFQWMAKPYVTVYTG	818	
CC						
DR	EMBL; AF06515; AAB87383.1; -.	QY	354	TAQARAVIREHEFYLS---KDQKI KKKKSQISSSESOKKRIKFDVLLTSTSEMINLSDA	409	
DR	EMBL; U08379; AAC50228.1; -.	Db	819	DKDSRAIIRENEFSFEDNAIKGKAFKWK-----REAQKFHVLLTSTVELTIDQA	870	
DR	Genew; HGNC:1918; CHD3.					
DR	MIM; 602120; -.	QY	410	VLKPIKCEMIVDEGHRLLNKDKSLFSSLTQYSSNHRILLTGTPLQNNLDELFLMHFLD	469	
DR	GO; GO:0004003; F:ATP dependent DNA helicase activity; TAS.	Db	871	ALGSIRWACLVDDEAHLKNNQSKFRVLNGYKIDHKLLTGTPLQNNLEELFHLNFLT	930	
DR	GO; GO:0003677; F:DNA binding; TAS.					
DR	GO; GO:0008270; F:zinc ion binding; TAS.	QY	470	AGKPSLSEPOEFKDIQNEQISRLHKLWLAPELLRRVKVKQVMKMPKPKELILRVDLSS	529	
DR	GO; GO:0007001; P:chromosome organization and biogenesis (sen. . .; TAS.	Db	931	PERFNNLEGFLSEFADISKEDQIKKLDLGLPHMLRLKADYFKNMPKTELIVRVELSP	990	
DR	GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.	QY	530	LQKEYYKAIFTNRYQVLTCKGGA-QISLNNTIMMELRKVCCHPYMLEGEPVHIDANEAPK	588	
DR	InterPro; IPR000953; Chromo.	Db	991	MOKKYKILTRNFALNSRGNGQVSLNIMMDLKKCNHPYLF----PVA--AMESPK	1044	
DR	InterPro; IPR001410; DEAD.					
DR	InterPro; IPR002464; DEAH_box.	QY	589	-----QLLESCKLQLLDKMMVKLKEQCHRVLIYTFQFQHMJLDLLEDYCTHKWQYE	639	
DR	InterPro; IPR001650; Helicase_C.	Db	1045	LPSGAYEGGALKSSGKMLLQKMLRKLKEQCHRVLIIFSQMTKMLDLLEDYDYGKYE	1104	
DR	InterPro; IPR000330; SNF2_N.	QY	640	RIDKVGGAERQIRIDRFNAKSNKFCFLSTTRAGGLGINLATATDTVIYDSQWNPHADL	699	
DR	InterPro; IPR001965; Znf_PHD.	Db	1105	RIDGGITGALRQEAIDRFNAPGAQCFCLLSTTRAGGLGINLATATDTVIIFDSQWNPNDI	1164	
DR	Pfam; PF00385; chromo; 1.	QY	700	CAMARAHRLGQTKVMYIELINRGCTIERRMQLTKKKMLVHLVGVK---LKTQINQEE	756	
DR	Pfam; PF00271; helicase_C; 1.	Db	1165	QAFSAHRIGQANKVMYIFVTRASVEERITQVAKRKWMLTHLVVRPGLSGKAGSKQOE	1224	
DR	Pfam; PF00628; PHD; 2.					
DR	Pfam; PF00176; SNF2_N; 1.	QY	757	LDIIIRYSGSKELFASDEDGAK---SGKIHYDDAIDKLDRLDLVEAEVSVDDSEENG	813	
DR	SMART; SM00298; CHROMO; 2.	Db	1225	LDLILKFGTEELFKDENGEGENKEEDSVIHYDNEAIAIRLLDRNQDATEDTDV--QNMNE	1282	
DR	SMART; SM00487; DEXDC; 1.					
DR	SMART; SM00490; HELIC; 1.					
DR	SMART; SM00249; PHD; 2.					
DR	PROSITE; PS00598; CHROMO_1; 1.					
DR	PROSITE; PS50013; CHROMO_2; 2.					
DR	PROSITE; PS00690; DEAH ATP HELICASE; 1.					
DR	PROSITE; PS01359; ZF_PHD_1; 2.					
DR	PROSITE; PS50016; ZF_PHD_2; 2.					
KW	Chromatin regulator; DNA-binding; ATP-binding; Helicase;					
KW	Nuclear protein; Repeat; Transcription regulation; Antigen;					
KW	Zinc-finger.					
FT	ZN_FING 379 426 PHD-TYPE 1.					
FT	ZN_FING 456 503 PHD-TYPE 2.					
FT	DOMAIN 494 594 CHROMO 1.					

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AF119716; AAD17276.1; -
 DR EMBL; AF003515; AAF49099.2; -
 DR EMBL; AY113368; AM29373.1; ALT_INIT.
 DR FlyBase; FBgn0013591; Mi-2.
 DR GO; GO:0005634; C:nucleus; ISS.
 DR GO; GO:0005700; C:polytene chromosome; IDA.
 DR GO; GO:0003682; F:chromatin binding; ISS.
 DR GO; GO:0004386; F:helicase activity; ISS.
 DR GO; GO:0016564; F:transcriptional repressor activity; IGI.
 DR GO; GO:0006333; P:chromatin assembly/disassembly; ISS.
 DR InterPro; IPR000953; Chromo.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002464; DEAH box.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR000330; Snf2_N.
 DR InterPro; IPR001965; Znf_PHD.
 DR Pfam; PF00385; chromo; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00628; PHD; 2.
 DR Pfam; PF00176; Snf2_N; 1.
 DR SMART; SM00298; CHROMO; 2.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELIC; 1.
 DR SMART; SM00249; PHD; 2.
 DR PROSITE; PS00598; CHROMO_1; FALSE_NEG.
 DR PROSITE; PS00013; CHROMO_2; 2.
 DR PROSITE; PS00690; DEAH ATP HELICASE; 1.
 DR PROSITE; PS01359; ZF PHD 1; 2.
 DR PROSITE; PS0016; ZF PHD 2; 2.
 KW DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat;
 KW Transcription regulation; Repressor; Zinc-finger.
 FT ZN_FING 377 424 PHD-TYPE 1.
 FT ZN_FING 437 484 PHD-TYPE 2.
 FT DOMAIN 488 566 CHROMO 1.
 FT DOMAIN 612 673 CHROMO 2.
 FT NP_BIND 755 762 ATP (POTENTIAL).
 FT SITE 875 878 DEAH BOX.
 FT DOMAIN 13 16 POLY-GLU.
 FT DOMAIN 70 76 POLY-LYS.
 FT DOMAIN 239 248 POLY-GLU.
 FT DOMAIN 1279 1287 POLY-GLU.
 FT DOMAIN 1672 1677 POLY-ASP.
 FT MUTAGEN 737 737 G->D; IN ALLELE MI-2-5; LARVAL LETHAL.
 FT CONFLICT 101 101 G -> A (IN REF. 1).
 SQ SEQUENCE 1982 AA; 224199 MW; ED8E256D1AD0AC2F CRC64;
 Query Match 28.5%; Score 2045.5; DB 1; Length 1982;
 Best Local Similarity 35.3%; Pred. No. 5 6e-96;
 Matches 513; Conservative 215; Mismatches 449; Indels 277; Gaps 43;
 QY 46 DAKENACQAGSTNLVSNCTCYAFHAKLVPLPKDASVNRWRCPEVSP--LNEIDKI 103
 DB 434 DEHQEFRCVCKGGELLCDCSPSAVHTFLNPLDITPDGDWRCPCRCPLTGKAEKI 493
 QY 104 LDCMRPTK-----SSEQSSDAEPKPIFVKYLVKWKGLSYLHCSWPE----- 148
 DB 494 ITRWRAQRNDGPFSTSGSKSNSR---VREYFIKHNNSYWHCEWPEVQLDVHHPLM 550
 QY 149 -KEFOKAY-----KSHRLKTRVNNFHRQMFNFNSDDFV-----AIRPE 188
 DB 551 IRSFQRYDWEPPKFEESLDEADTRYKRIQRHKDKVGMKANDAEVLEERYFKGVKPE 610
 QY 189 WTTVDRLACR-BEDGEYLEYLVKYSDECYWESE--SDISTFQNEIQRFKDV----- 240
 DB 611 WLIVQVRVNHRTARDGSTWYLVKWRPELYDKSTWEEEGDDIQGLRQAIDYQDLRAVCTS 670
 QY 241 ---NSRTRSK-----DVDHKNRPD-----FQQPDHTPEFLKGL---L 273

DB 671 ETTQSRSKSKGRKSKLKVDEDEDRPVKHYTPPEKFTTDLKKYEQPAFLEGTGQWL 730
 QY 274 HPYQLEGINFLRPSWSKQTHVILADEMGLGKTIQSIALLASFEENLI--PHLVITAPST 331
 DB 731 HPYQIEGINWLRYSWGQGITILADEMGLGKTIQTVTFLYSLYKEGHCRGPPLVAPLST 790
 QY 332 LBNWEREFATWAPQWVVMYFGTAQARAVIREHEFYLSKDQKKIKKSKSGQISSSKQKR 391
 DB 791 LVNWEREFELWAPDFYCITYIGDKDSRAVIRENE--LSFEGAIRGSKVSRL----RTQ 844
 QY 392 IKFVDLLTSYEMINLDSAVLKPIKWECMIVDEGHRKKNKDSKLPSSLFQYSNHRILLTG 451
 DB 845 YKFNVLTSYELISMDAACLSIDWAVLVVDEAHRKLSNQSFKFFILNSYTTIAYKLLTG 904
 QY 452 TPLQNNLDELFMHFLDAGKPGSLSEFQEFKDIQBEQISRLHKMLAPHLRLRRVKDV 511
 DB 905 TPLQNNLEELPHLLNFLSRDKFNDLQAFQGEFADVSKEEQVKRLHEMLGPHMLRLKTDV 964
 QY 512 MKDMPKXELILRVDLSSLQKEYKAIPTRYQVL-TKKGGAQISLNNIMMELRKVCCHP 570
 DB 965 LKNMPSKSEFIVRVLSAMQKKFYKFIILTNYEALNSKSGGSCSLINIMMDLKKCCNHP 1024
 QY 571 YMLEGVEPVIHDANE-----FKOLLESGLQLLDKMWKVLKEQGHRLVLYTQFO 621
 DB 1025 YLFPSA-----ABEATTAAGLYEINSLTKAAGKLVLLSKMLKQKQNHVLIQSMT 1078
 QY 622 HMLDLLEDYCTHKKWYERIDKVGGAERQIRIDRFNAKNSNKFCLLSTRAGGGINLA 681
 DB 1079 KWLIDILEDFLEGEQKYERIDGGITGTLRQEAIDRFNAPGAQFVFLSTRAGGGINLA 1138
 QY 682 TADTVIIYDSWNPHADLOAMARHLGQTKWMIYRLINRGTIIEERMQLTKKQVLEH 741
 DB 1139 TADTVIIYDSWNPHNDIQAFSRAHRIGQANKWMIYRFVTRNSVEERTVQAKRKMLTH 1198
 QY 742 LVV-----GKLTQNIQOEELDDIIRYGSKELFASEDEAGSGKGIHYDDAIDKLDRD 796
 DB 1199 LVVRPGMG--KGANFTKQELDDILRFGTEDLFDKEDKEBA---IHDDDKAVABELDT 1252
 QY 797 LVAEVSVDDDEENGFLKAFKVFANFEYIDENEAALAEQARVAESKSSAGNSDRASYE 856
 DB 1253 NRGIEE---KESWANEYLSFKVASYATKEEBEETEI-----IKQDAENSDPA-YVW 1302
 QY 857 ELLKDKFELHQAELNALGKGRSRKQL-----VSIEDDLAILEDVSSDGESEY 907
 DB 1303 KLLRHHYEQHOEDVGRSLGKGRVKQVNYTDGGVVAADTTTRDD--SNMQDNGSEYNSEYS 1361
 QY 908 AES-TDGEAAGQGVOTG--RRPYR- KGRDNLPTPLMEGEGRSFVLGNQSQRATVQ 963
 DB 1362 AGSDEGDDDDFDQNGAERKAKRLERRDRPLPLIARVGGNTEVLGFNARQKSFNL 1421
 QY 964 TLMRYG-----AGNFDWKEFVRLKQKTFEEINEYGILFLKHIAEIEDENSPFSDGVPK 1018
 DB 1422 AIMRYGMPQDAFNQW--LVRLRGKSERNFKAVSLFMRHLCPGADNAETFADGVP 1479
 QY 1019 EGLRIEDVLVRIALLILVQEKVQFVEDHPGKVPFSPRILERPPLGURSKIKWEHDKIMI 1078
 DB 1480 EGLSRQHVLTIRIGVMSLRKKVQEPHINGYYSMPELIKKPCPVRS----- 1526
 QY 1079 RAVLKHGIVRWQAI VDDKELGIQELI CKELNFPHTLSLAAEQALQGGQSGSGSNPQAOT 1138
 DB 1527 --ALKQ-----DVAALAEAPTGGNVDSKAT 1550
 QY 1139 NONPGSVITGNNNASADGAQVNSMFYRDMQRRLLVEFKVRVLLLEKAMNYEAEYYGL 1198
 DB 1551 SNS-----VTPATSAAPSAPASEKEDKDS-----EKEKDTSAEK---- 1589
 QY 1199 GSSSIPTEEPAEAPKPIADTVGVSPFIEVDDMLGLPKTDPTITSEEINGAAVDNN-QARV 1257
 DB 1590 ----SEVQEQEAE-----EDKKPGDVQENPV--BEAAGDTKPSDAEYKT 1629
 QY 1258 EIAQHYNQMKLLDENARESVAQVYNNQPPSTKYNESFRALKSINGNINTILSITSDQSK 1317

```
Db 1630 EVAK-----TEPKETKQPEVKBPK-----TBEKEK 1656
Qy 1318 SHEDDTKP-----DLNNVEMKDTAEETKPLRGVVDLNVVGEENIAEASGVVDV 1367
Db 1657 EKVDKKPIPTTVIDDDDDVMIVKEDGELEKPSASSPKQKAVAAA--TSAATGATGK 1714
Qy 1368 KMEEAKEEEKPKNM 1381
Db 1715 GAEDSLEVLKRPKM 1728

RESULT 4
CHD3 CAEEL
ID CHD3 CAEEL STANDARD; PRT; 1787 AA.
AC Q22516; Q18794;
DT 15-JUL-1999 (Rel. 38, Created)
DT 25-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chromodomain helicase-DNA-binding protein 3 homolog (CHD-3).
GN CHD-3 OR T14G8.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20530482; PubMed=11076750;
RA von Zellewsky T., Palladino F., Brunschwig K., Tobler H., Hajnal A.,
RA Mueller F.;
RT "The C. elegans Mi-2 chromatin-remodelling proteins function in vulval
RT cell fate determination.";
RN Development 127:5277-5284 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Matthews P., McMurray A.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: Chromatin-remodelling protein that function in vulval
CC cell fate determination.
CC -I- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -I- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC -I- SIMILARITY: Contains 2 PHD-type zinc fingers.
CC -I- SIMILARITY: Contains 2 chromo domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF308444; AAG29837.1; -.
CC EMBL; Z67884; CAA91810.1; -.
CC EMBL; Z67881; CAA91810.1; JOINED.
CC EMBL; Z67881; CAA91798.1; -.
CC EMBL; Z67884; CAA91798.1; JOINED.
CC PIR; T20160; T20160.
CC
CC WormPep; T14G8.1; CE03657.
CC InterPro; IPR000953; Chromo.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002464; DEAD box.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR000330; SNF2_N.
CC InterPro; IPR001965; Znf_PHD.
CC InterPro; IPR001841; Znf ring.
CC Pfam; PF00385; chromo; 1.
CC Pfam; PF00271; helicase_C; 1.
CC Pfam; PF00628; PHD; 2.
CC Pfam; PF00176; SNF2_N; 1.
CC SMART; SM00298; CHROMO; 2.
CC SMART; SM00487; DEXDC; 1.
CC SMART; SM00490; HELIC; 1.

DR SMART; SM00249; PHD; 2.
DR SMART; SM00184; RING; 2.
DR PROSITE; PS00598; CHROMO 1; FALSE_NEG.
DR PROSITE; PS00013; CHROMO 2; 1.
DR PROSITE; PS00890; DEAH_ATP_HELICASE; 1.
DR PROSITE; PS01359; ZF_PHD 1; 2.
DR PROSITE; PS00016; ZF_PHD 2; 2.
KW Chromatin regulator; Nuclear protein; Repeat; Helicase; DNA-binding;
KW ATP-binding; Zinc-finger.
FT DOMAIN 59 62 POLY-LYS.
FT ZN_FING 265 312 PHD-TYPE 1.
FT ZN_FING 328 375 PHD-TYPE 2.
FT DOMAIN 373 476 CHROMO 1.
FT DOMAIN 501 583 CHROMO 2.
FT DOMAIN 1287 1291 POLY-ARG.
FT NP_BIND 641 648 ATP (POTENTIAL).
FT SITE 763 766 DEAH BOX.
SQ SEQUENCE 1787 AA; 205254 MW; 1EFC1FFECB59740 CRC64;

Query Match 28.3%; Score 2032; DB 1; Length 1787;
Best Local Similarity 33.5%; Pred. No. 2.3e-95;
Matches 511; Conservative 229; Mismatches 438; Indels 346; Gaps 45;

Qy 19 NLDDSDDDDF-VPKKDRTFEQVEAIVRTDAKNA--COACGESTNLVSCNTCTYAFHAKC 75
Db 295 NMEQPPEGDWSCPHCEHGPDLIVVEEPAKANMDYCRICKETSNILLCDTCFSSYHAYC 354
Qy 76 LYPPLKDA5VENWRCPECV--SPLNEIDKILDC-----EMRPTKSEQSSD----- 120
Db 355 IDPPIETPEGWSGPCRIIPPAQRIEKILSWRWEKISYPPLECKEKEEASKDDVFLK 414
Qy 121 --AEPKPIFKOYLKWKGLSYLHGSVWPEKEFKAYKSNHRL---KTRVNN---PHRQM 172
Db 415 PPKMEPREREFFVKWKYLAWQCEWLSLMDVYFTALVRMYWRKVDSENPIFEEST 474
Qy 173 ESNFNSDD-----FVAIRPWTTVDRIL-----ACREDEGELEVLVKYKELSYDE 218
Db 475 LSRHSHDHPYKLRERFYQYGVKPEWQIHRINHLSYAKSQD-----YLVWKELSYEH 530
Qy 219 CYWE-SESDISFQNEIQRF-----KDVNSRTRSK-----DVDHKK 254
Db 531 ATWERDDTDIANYEDAILIKYWHHRERMLNDEVPRNVQKMIQAKQREKGLPKGEVTSRR 590
Qy 255 NRPD-----PQQDHTPEFLK---GLLHPYQLEGLNPLFRFSWKSQTHVILADEMGLKTTQ 307
Db 591 KKREKIDILKYEVPDFIFSETGNGNLHPVQLSGINLWRCWNGSNGTDAILLADEMGLKTVQ 650
Qy 308 STALLASLPEENLI--PHLVIAPLSTLRNWEREFATWAPQMVNVYFGTAQARAVIREHE 365
Db 651 SLTFLYTLMEGHTKGPFLIAAPLSTIINWEREAELWCDFYVTVYVGDRESRMVIREHE 710
Qy 366 FYL-----SKDQKKIKKKSGQISSKQKRIKFDVLLTSEYMINLDSAVLKPIKWCMTV 421
Db 711 FSVFDGAVRGGPKVKSKIT-----LENLKFHVLTLTSEYCNMDKALLSSIDWAALVV 762
Qy 422 DECHRLKNKDKSLFSLSTQYSSNHRILLTGTPLQNNLDELFLMLHFLDAGKFSLEFQER 481
Db 763 DEAHRLKNNQSTFFKNLREYNIQYRVLLTGTPLQNNLEELFLLNPLAPDRFNQLESFTA 822
Qy 482 EFKDINQEQISRLRHKLAPHLRLRVKVKDKMDKMPKELILRLVDLSLQKEYKAIFTR 541
Db 823 EFSEISKEDQIEKLNHLNLPMLRLKADVLTKGMPSPKQELIVRVELSAMQKKYKKNILTR 882
Qy 542 NYQVL--TKKGAQISINNTMELRKVCCHPYM-----LEGVPEVTHDANEAFKQLLESCEG 595
Db 883 NFDALNVKNGGTQMSLINIIMELKCCNHPYLFMKACLEA--PKLKNMGVEGSALLKNAG 940
Qy 596 KLQLLDKMMVKLKEQGHRRVLIYTFQFQHMLDLLEDYCTHKKQYERIDGVKGGAEQIRID 655
Db 941 KVVLLQKMLRKLKDGGRVLIIFSQMTMLDILEDFCDEVGKYKVERIDGSIQQRQDAID 1000
Qy 656 RFNAKNSKFCPLLSTRAGGLGINLATDTVIIYSDSNPHADLQAMARHLRGQTNKUM 715
```



```
Db 1001 RYNAPGAKQVFLSTRAAGLGINLATADTVIIYSDMNPNDIQAFSRAHRLGQKHVM 1060
QY 716 IYRLNWRGTEIERMMQITKKQVLEHLV-----GKLTQNIQNEELDDIIRVGSKELPAS 771
Db 1061 IYRFVTGSGVEERITSVAKKCMLLTHLVVRAGLGKDGKSGMSKTELDVLRWGTBELFXE 1120
QY 772 ED-----DEAGSGK-----LHYDDAIDKLLDRDLVEARVSVDDREE-----NGFLK 815
Db 1121 EAPVGAOGEGTSSKKPNQEIVWDAAVDFLDRN---KEEGDQGGKKHWNTEYUS 1177
QY 816 AFKVFANPEYIDENEAALAEQRAVAESKSSAGNSDRASYEELLKDKFELHQAEELNALG 875
Db 1178 SPKVATY---NTKEADDAADDEDETEVIKGETEEDQPNYWEKLLKHVYQDQETELQKLG 1234
QY 876 KKRKSRKQLVSIHED---DLAGLEDVSSDGESEYAEASTDGERAGOGVOTGRPRVYRKOR 932
Db 1235 KGRVRVRQVNYASENNQDWSAQNNQBEDDQSGYSDNGELLQTDDEYEEERRRRREER 1294
QY 933 DNLEPTPLMEGEGRSFVLGFGNOSORAI FVOTLMRYGAGNFD-----WKEFPRLKQKTF 987
Db 1295 SEKLP-PLLAKVNGQTEVLGFNPRQKAFVNAVWRMGMPQDITQSSWQ--VRDLRNKSE 1351
QY 988 EETNEYGILFLKHIABEIDENSPTFSGDGVPKEGLRIEDVLRIALLILVQEKV---KF- 1042
Db 1352 KVFKAYSSLFMRHLCEPVDNSDFWDGVPREGLNQAVLSRIGLSILRKVKVQEFKEN 1411
QY 1043 ----- 1042
Db 1412 GEWSMPETREKMLATAAQASVSNLPGMIKIKEPIDIDETPMVDQSDNITKTEELASEVK 1471
QY 1043 VEDHPCKVPFPSRI-----LERPFLRSK---TWKEHDKIMIRAVLKHG 1085
Db 1472 VEEPKAPRLPYKFNICDGYTELHSLWNEEKVARGNKEYEIIWHRHDFWLLAAVAVYG 1531
QY 1086 YGRQWQIVDDKIGLIGELICKELNFPHISLSAAEQAGLQCGNGSGSNPGCAOTNONGPSV 1145
Db 1532 YGRYQ-----INFODI---MNDPKFSIV-----NEFPKGTGADPATN 1565
QY 1146 ITGNNASADGAGVSMFYRDMQRLVFEVKRVLLLBKAMNYEYAEYVYGLGGSSSTP 1205
Db 1566 FADVKV-----KFLARRFKLLBQSL----- 1585
QY 1206 TEEPEAEPKIADTVGSFIEVDDEMLDGLPKPTPTTSEEIMGAVDNNQARV-----EIAQ 1261
Db 1586 -----VIEQLRRAAHINKQOOSPDQVQGLAQ 1611
QY 1262 HYNQMCKLLDEN---ARESVQAVNNQPPSTKWNESFRALKSINGNINILTSITSDQSKS 1318
Db 1612 HFSELEHTADAHVNIARES-----NNG-----NRNANAILHKCLAQLD 1650
QY 1319 HEDDTKPDLLNNVEMKDTAETKPL 1342
Db 1651 LLSLDLKTDV--ARLPATISQVRV 1672

RESULT 5
CHD3_DROME STANDARD; PRT; 892 AA.
ID O16102; Q8SVJ8; Q9VW23;
AC 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
OS Chromodomain helicase-DNA-binding protein 3.
GN CHD3 OR CG9594.
OC Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=Berkley.
RX MEDLINE=20196006; PubMed=10731132;
```

```
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George K.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aqbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Botchan M.R., Bouch J., Brokstein P., Brottier P.,
RA Burtis K.C., Bussey D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischman W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasako P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RP REVISIONS.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J.R., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN [4]
RP SEQUENCE OF 1-812 FROM N.A.
RX MEDLINE=97470991; PubMed=9326634;
RA Woodage T., Basrai M.A., Baxeianis A.D., Hieter P., Collins F.S.;
RT "Characterization of the CHD family of proteins.";
CC Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997).
CC -!- FUNCTION: Possible transcription activator (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC -!- SIMILARITY: Contains 2 chromo domains.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to a stop
CC codon in position 628 and a frameshift in position 720.
```


DR Genew; HGNC:1917; CHD2.
DR MIM; 602119; --
DR GO; GO:0004003; F:ATP dependent DNA helicase activity; TAS.
DR GO; GO:0003677; F:DNA binding; TAS.
DR GO; GO:0007001; P:Chromosome organization and biogenesis (sen. .; TAS.
DR GO; GO:0006357; P:regulation of transcription from Pol II pro. .; TAS.
DR InterPro; IPR000953; Chromo.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR003330; SNF2_N.
DR Pfam; PF00385; chromo; 2.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00298; CHROMO; 2.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00598; CHROMO 1; 2.
DR PROSITE; PS00113; CHROMO 2; 2.
DR PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE NEG.
KW Hydrolase; Helicase; Nuclear protein; ATP-binding; DNA-binding;
KW Repeat.
FT DOMAIN 1 74 SER-RICH.
FT DOMAIN 121 139 SER-RICH.
FT DOMAIN 261 353 CHROMO 1.
FT DOMAIN 378 456 CHROMO 2.
FT NP_BIND 509 516 ATP (POTENTIAL).
FT SITE 617 620 DEAH BOX.
SQ SEQUENCE 1739 AA; 200560 MW; D62089C25EBA440F CRC64;

Query Match 22.6%; Score 1621.5; DB 1; Length 1739;
Best Local Similarity 30.7%; Pred. No. 1.5e-74;
Matches 453; Conservative 229; Mismatches 493; Indels 299; Gaps 43;

QY 11 RSDRPVNLDDSDDDDFVFK-----DRTFQVEAIVRTDAKENACQACG 56
DB 199 RGKRRKQDSDDDDDDEAPKQTRRAAKNVSYKEDDFE-----TDS--DDLIEWTG 250
QY 57 ESTNLVNCNTVAFHAKCLVPLKDAVENRCPECVSPLEIDKILDCENRPTKSEBQ 116
DB 251 EGVDEQDQDQNS-----IEKVLD--SRLGKKGAT 277
QY 117 GSS-----DAEPKPIFVK-----QYLVKWKLSVLHCSWVPEKEFO----KAYS 157
DB 278 GASTTVYAEANGDPGDFTEKDEGEIQLIKWKGWSYHSTWSEESLQOQKVGLKK 337
QY 158 NHRLLKTRVNFNR-----QWESFNNSDDFVAIRPEWTTVDRIACREE----- 201
DB 338 LENFKKDEDEIKQWLGVKSPDEVEYFNCQELASELNKQYQIVERVIAVKTSGTLGQTD 397
QY 202 -----DGELEYLVKYKLSYDECYWESESDI-STFQNEIQRFKDVNSR-----TRR 246
DB 398 FPAHSRKPAAPSNEPEYLCKWMLPYSECSWEDEALIGKFKQNCIDFSHRNNSKTIPTRE 457
QY 247 SKVDHKNRPRDQFDHTPEFLG---LHPYQLEGLNLFPSKQTHVILADEMGLG 303
DB 458 CKAL--KORPR-FVALKKQPAYLGGNLELDYQLEGLNWLASHCSWKNVSLADEMGLG 514
QY 304 KTIQSIALLASLFEENLI--PHLVAPLSTLNWEREFATWAPQMVNVYFGTAQARAVI 361
DB 515 KTIQTSIFLSYLFHQQLGPFLLVPLSTLSWQREFEIMAPEINVVYIGDLNRSNTI 574
QY 362 REHEFYLSKQDKKIKKKKSGQISSKSKQRIKFPDVLITSYEMINLDSAVLKPWKECMIV 421
DB 575 REYEMI-----HSQTKRLKFNALITTYEILLKDKTVLGSINNAFLGV 616
QY 422 DEGRLLKNVDSKLFSSLTQYSSNHRILLGTPLONLDELPMHPDLADGKGSLEBQ 481
DB 617 DEAHRLKNDSDLYKTLIDFKSNRLLITGLQNSKLWELSHLFIIMEKFEFDEFE 676
QY 482 EFKDINQEEQISRLHKLAPHLLRRVVKDKVMDMPKPELILRLVDLSLQKQYKAIPT 541
DB 677 DHGK-GRENGYSLHKVLEPFLLRRVKKDVKEKSLPAKVEQILRVENSALQKQYKWLTR 735

QY 542 NYQVLTG-KGGAQISLNNIMMELRKVCCHPYMLEGVEP-VIHDANEAFKQLLESCKQLQ 599
DB 736 NYKALAKTRGSGTGFNLVIMVELKCCNHCYLLIPEENRENGOEILLISIRSGKIL 795
QY 600 LDKMVKLEQCHRVLIYTFQHMLEDLECYTHKKWOYERIDGKVGGAERQIRIDRENA 659
DB 796 LDKLLTRLRGRNVLIIFSQVMRLDILAELYTIKHYFFQRLDGSIKGEIRKQALDHFA 855
QY 660 KNSKPCFLSTRAGGLGINIATDVIIYSDWNPHADLQAMARAHLRGQTNKVMYRL 719
DB 856 DGSDFCELLSTRAGGLINLASADTVVIFSDWNPNQDLQARAHRTGKQKQVNIYRL 915
QY 720 INRGTIERRMQLTKKQWVLEHLV-----GKLTQNTNQELEDIIRVG 764
DB 916 VTGTFVEEIERAKKQWLDHLVITQRMDTGRTILENNSGRSNPNFPKKELTAILKFG 975
QY 765 SKELFASDDDEAGSKGIHYDDAIDLKLLDRDLVEAEVSDVDEEENFLKAFKVANFEY 824
DB 976 AEDLFKELEGESEF-----QEMDIDEIL--RLAETRENEVSTSATDELLSQFVANFAT 1028
QY 825 IDENEAALAEQARVAESKSSAGNSDRASYWEELL-----KDFELHQAELNALGK 876
DB 1029 MEDEE--ELEER-----PHKD-----WDEIIPERQKKVEEERQKLEIYMLPR 1072
QY 877 RKRSEKQLVSIEDDDLAGLEDVSSDGDSEYAE-STDGEAAGQGVOTGRRPYRRKGRDL 935
DB 1073 IRSSTK-----AQTNDSDSDTESKROAQSSASESTEDSDDDKKPKRR----- 1117
QY 936 EPTPLMEGEGRSFR---VLGFNQORAFVOTLMRYGAGNFDWKEFVPR---LKOTFEE 989
DB 1118 -----GRPSRVKDLVEGFTDAIRFIIKAYKFGI-PLERLECLARDDELVDKSVAD 1169
QY 990 INEXGILF-----LKHIAEIDENSPTPSDGVKPEGLRIEDVLVRIALLILVQEKVF 1042
DB 1170 LKRLGELTHNSCVSAMQSYEBQKENASEGKPGKRRGPTIKISGVQNVKSIHQHEBF 1229
QY 1043 VEDHPGKVPFSR-----ILERFPLRSKGTWKEHDKIMIRAVLKHGYSQWQAIVDKE 1097
DB 1230 EMLKHSIPDPEEKKKCYCLTCRVAAHPDVEGWEDDSRLLLGIYEHGYGNWELIKTDP 1289
QY 1098 LGIOELIC-----KELNFPHISLSAAEAGLQONGSGSGNPGAQTNQNPQSVIT 1147
DB 1290 LKLTDKILPVETDKPKQKQIOTRADYLLKLLRKGLEK-----GAVT 1332
QY 1148 GNNASADGAQVNSMFFYRDMQRRLVEFVKRLLLEKAMNYEYAEYVYGLGSSSIPTE 1207
DB 1333 GGEAK-----LKKRKPRVKKENKVPRLKEEHGIELSPRSHD 1370
QY 1208 EPEAPKPIADTVGSFIEVDDEMLDGLPKTDPITSEEINGAAMDNNQARVEIAQHNQMC 1267
DB 1371 NPSEGEVKD-----DGLEKS-PMKKKKQKKENK-----KQMS 1406
QY 1268 KLLD-ENARESVOAVVNNPPSTKVNESFRALKSINGNITILSITSDQSKSHEDDKPD 1326
DB 1407 SRKQEGEKERRKSKDKEKPKSGDAKSSSKSKSGQGVPHITAGSEVPVIGEDED---D 1463
QY 1327 LNNVEMKDTABETPLRGVVDLVNVEGEENIAE 1360
DB 1464 LDQETFSICKERMMPVKKALKQLDKPKGLNVQE 1497

RESULT 7
CHD1_MOUSE
ID CHD1_MOUSE STANDARD; PRT; 1711 AA.
AC P40201.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Chromodomain-helicase-DNA-binding protein 1 (CHD-1).
GN CHD1 OR CHD-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.


```
Db 1287 LPDDPKKPOAKLOTRADYLIKILSLDLAKREAOQLCAGGS----- 1329
QY 1150 NNASADGAQVNSMFYRDQWRLBFVKRVLLEKAMNVEYARYGLGGSSIPTPEP 1209
Db 1330 -----KRRKTRAKSKAM-----KSKVKEEI--KSDSSPLPSEKS 1363
QY 1210 EAEPKIADTVGVSFIEVDDEMLDGLPKTDPITSEIM-----GNAVDNNQARVEI 1259
Db 1364 D-----EDDDKLANDSKPESKORSKSVSDAPVHITASGEFVPFAESEBEL 1409
QY 1260 AQHYNQMKC 1268
Db 1410 DQKTSICK 1418

RESULT 8
CHD1 HUMAN
ID CHD1 HUMAN STANDARD; PRT; 1709 AA.
AC O1466;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Chromodomain-helicase-DNA-binding protein 1 (CHD-1).
GN CHD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97470991; PubMed=9326634;
RA Woodage T., Basrai M.A., Baxevanis A.D., Hieter P., Collins F.S.;
RT "Characterization of the CHD family of proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997).
CC -!- FUNCTION: Sequence-selective DNA-binding protein. Could play an
CC important role in gene regulation.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC -!- SIMILARITY: Contains 2 chromo domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF006513; AAB87381.1; -.
DR HSSP; P23197; IAP0.
DR Genew; HGNC:1915; CHD1.
DR MIM; 602118; -.
DR GO; GO:0004003; F:ATP dependent DNA helicase activity; TAS.
DR GO; GO:0007001; P:Chromosome organization and biogenesis (sen. .; TAS.
DR GO; GO:0006357; P:regulation of transcription from Pol II pro. .; TAS.
DR InterPro; IPR000953; Chromo.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00385; chromo; 2.
DR Pfam; PF00371; helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00598; CHROMO_1; 2.
DR PROSITE; PS00133; CHROMO_2; 2.
DR PROSITE; PS00690; DEAD ATP HELICASE; FALSE NEG.
KW Helicase; Helicase; Nuclear protein; ATP-binding; DNA-binding;
KW Repeat.
FT DOMAIN 1 70 SER-RICH.
FT DOMAIN 117 137 SER-RICH.
FT DOMAIN 272 364 CHROMO 1.
```

```
FT DOMAIN 389 452 CHROMO 2.
FT NP_BIND 506 513 ATP (POTENTIAL).
FT SITE 614 617 DEAH BOX.
FT DOMAIN 1628 1644 3 X 5 AA REPEATS OF H-S-D-H-R.
FT REPEAT 1628 1632 1.
FT REPEAT 1634 1638 2.
FT REPEAT 1640 1644 3.
FT SEQUENCE 1709 AA; 196517 MW; 416409C913D6A935 CRC64;

Query Match 21.7%; Score 1562.5; DB 1; Length 1709;
Best Local Similarity 32.4%; Pred. No. 1.5e-71;
Matches 410; Conservative 220; Mismatches 384; Indels 251; Gaps 41;

QY 9 RIRSDRPVNLDDDDDFVPKDKRTFEQVEAIVRTDAKE-----NACQACGSEST 59
Db 204 KILGQKKRQIDSSEDDDEEDYDNDKRSRRQATVNVSYKDEEMKTDSDLLLEVCGSD- 262
QY 60 NLVSCNTCTYAFHAKCLVPLKDAVENWRCPCEVPLNEIDKILDCSM-----RPTKS 113
Db 263 -----VPOEEBEFET-----IERPMDCRIGKGATATT 293
QY 114 SEQSSDAEPKPIFVK-----QYLVKWKGLSYLHCSWVPEKFKQKAYKSNHRUKRVN 166
Db 294 IYAVEADGPNAGFEKNKEPGEIQYLIKWKQWSHHNTWETEELK---QQNVKMGKKLD 350
QY 167 NFHRQ-----MESFNSEDDDFVAIRPEWTTVDRIILACREED---CELYL 208
Db 351 NYKKKQDQETKRWLNKNASPEVYNNCQQLDHLHQYQIVGRIIAHSNQKSAAGYPDY 410
QY 209 VKYKELSYDECVWESEDIS--TFQNEIQREKDVN--SRTRRSKDVN--HKNRPDQFQDHT 265
Db 411 CKWQGLPYSECSWEDGALISKKQFACIDEYFSRNSQKTTTPKDCVKLKQRP--FVALKKQ 469
QY 266 PEFKLG-----LLHPYQLEGLNFRFSWSKQTHVLADEMGLGKTIQSIALLASFEENLI 321
Db 470 PSYIGHGHEGLRDYQLNGLNLAHSAWCKNSCILADEMGLGKTIQTSIFNLYPFHEQL 529
QY 322 --PHLVIAPLTLNRWEREPATWAPQNVVYFQTAQARAVIREHEFVYLSKDKKIKKK 379
Db 530 YGPELLVPLSTLTSWQREIQTWASQMNNAVYLGDISNRNMIRTHEW----- 576
QY 380 SGQISSESKQRIKFDVLLTSYEMNLDSAVLKPIKWCMIVDEGHRLNKDSDLFSSIT 439
Db 577 -----THQTKRLKFNILLTYIILLKDKAFGLGNWAFIGVDEAHRUNDDSLLYKTLI 631
QY 440 QYSSNHRILLTGTPQLNNLDELFLMHFADAGKFGSLEEFQEFKDIQNEEQISRLHKML 499
Db 632 DFKSNHRLITGTPLQNSLKELWSLHHPIMEKFTSSWEDFEEHCK--GREYGYASLHKEL 690
QY 500 APHLIRRVKQVMDMPKELIILRVDLSSLOKEYYKAFITRNQVLTG--KGAQISLNN 558
Db 691 EPFLIRRVKQVMDVEKSLPAKVEQILRMESALQKQYKWLTRNYKALSKGSGSTSGFLN 750
QY 559 IMWELRVKCCHPYMLEGVE--FVIDHANEAFKOLLESCGKLOLLDKMVKLEQGHRLVLI 617
Db 751 IMWELKCCNHCYLLKPPDNNFYNKQALQHLIRSSGKILLDKLLLRERGRNLIF 810
QY 618 TQFQMLDLLEDYCTHKKQYERIDGKVGGAERQIRIDRFNAKSNKFCFLSTRAGGLG 677
Db 811 SQVMRLDILAEYLYKRYQFPQRLDGSIKGELRQALDHFNAEGSEDFCLLSTRAGGLG 870
QY 678 INLATADTVIYDSQWNPADLOAMARHRLGQTNKVMYIRLNGRTTEERMMQTKKMM 737
Db 871 INLASADTVIYFDSQWNPQNDLOAARHRIQOKKQVNIYLVTKSGSVEEDILERAKKMM 930
QY 738 VLEHLVGVKLTQN-----INQELDDIIRYGSKELFASEDDACKSGKI 782
Db 931 VLDHLVIQRMDDTGTGTVLHTGSAPSSSTPPFNKEELUSALIKFGAEELFKPEGEQEP--- 987
QY 783 HYDDAAIDKLDRLDIVEAEE---VSDDEEENGFLKAFKVFANFEYIDENEAAALQAORVA 839
Db 988 --QEMWDIDELIKRAETHENEFGPLTVGDE---LLSQFKVANFNSWMDDD--LELEPER-- 1038
```

```
QY 840 ABSKSSAGNSDRASYWEEL-----LKDKFELHQAEELNALGKRRKSRKQLVSTIEDD 891
Db 1039 -NSKN-----WEEIIPEDQRLEERQKELEIYMLPRMNC AKQ----- 1079
QY 892 LAGLEVSDSGDESFAES-----TDGEAAGGVQTRPPYRKGRKRDNLPTPLMEGEGR 946
Db 1080 -----ISFNGSEGRSRSSRYSGSDS-----ISEGKRP-KRGRPRPTIPRE----- 1121
QY 947 SFRVLGFSQRAIFVQTLRMRYCAGNDFWKEFVPR---LKQTFPEINEYGLFLKHIAE 1003
Db 1122 --NIKGFSDAEIRRFYKFKFG-GPLERLDAIARDAELVDKSEYDRLRLGELVHNGCTK 1178
QY 1004 EIDENS-----PTFSDGVPKEGLRIEDLVRIALAILVQEKVKFVEDHPG 1048
Db 1179 ALKDSSTGRTGRLGKVGKPTF-----RISGVQVNAKLVISHEEL--IPLHKS 1227
QY 1049 KPVFPRIILERPPLRSGKI-----WKEHDKIMIRAVLKHGYGRQWQIAVDDK 1096
Db 1228 IPSDPE--ER---KQYTIPTCHTKAAHFDIDWGEDDSNLLIGIYEYGGSWEMIKMDP 1280
QY 1097 ELGIQELI-----CKELN-----FPHISLSAAEQAGLQGGNGSGSNPGAQTNQ 1140
Db 1281 DLSLTHKILPDDPKPQAKQLQTRADYLIKLSRLDLAKKEALSGAGSKRRKARAKNKK 1340
QY 1141 NPGSV 1145
Db 1341 AMKSI 1345

RESULT 9
CHDB HUMAN
ID CHDB HUMAN STANDARD; PRT: 2004 AA.
AC Q9HCF8; Q8N3Z9; Q8NCV4; Q8TR9; Q96F26;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Chromodomain-helicase-DNA-binding protein HELSNF1 (Helicase with SNF2
  domain 1) (Fragment).
GN HELSNF1 OR KIAA1564.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22158633; PubMed=12168954;
RA Nagase T., Kikuno R., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT "Construction of expression-ready cDNA clones for KIAA genes: manual
  curation of 330 KIAA cDNA clones.";
RL DNA Res. 9:99-106(2002).
RN [2]
RP SEQUENCE OF 588-2004 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirotsawa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
  XVIII. The complete sequences of 100 new cDNA clones from brain which
  code for large proteins in vitro.";
RL DNA Res. 7:273-281(2000).
RN [3]
RP SEQUENCE OF 1554-2004 FROM N.A.
RC TISSUE=Brain, and Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeng B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
```

```
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madañ A., Rodrigues S., Sanchez A.,
RA Whiting M., Madañ A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
  human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 1768-2004 FROM N.A.
RC TISSUE=Lymph node;
RA Duesterhoef A., Lauber J., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Probable transcription regulator.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC -!- SIMILARITY: Contains 2 chromo domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB046784; BABI3390.2; -.
DR EMBL; BC011695; AAH11695.1; ALT_INIT.
DR EMBL; BC025964; AAH25964.1; -.
DR EMBL; BC036920; AAH36920.1; -.
DR EMBL; AL834524; CAD39180.1; -.
DR Genew; HGNC:20153; HELSNF1.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0030528; F:transcription regulator activity; NAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR InterPro; IPR006576; BRK.
DR InterPro; IPR000953; Chromo.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00592; BRK; 2.
DR SMART; SM00298; CHROMO; 2.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00598; CHROMO_1; FALSE_NEG.
DR PROSITE; PS50013; CHROMO_2; 1.
DR PROSITE; PS00690; DEAD_ATP_HELICASE; FALSE_NEG.
KW Transcription regulation; Hydrolyase; Helicase; Chromatin regulator;
  Nuclear protein; ATP-binding; DNA-binding; Repeat.
FT NON_TER 1
FT DOMAIN 65 132 CHROMO 1.
FT DOMAIN 147 213 CHROMO 2.
FT NP_BIND 259 266 ATP (POTENTIAL).
FT SITE 371 374 DEAD_BOX.
FT DOMAIN 1492 1521 SER-RICH.
FT DOMAIN 1916 1931 HIS-RICH.
FT DOMAIN 1962 2004 ASP-RICH.
FT CONFLICT 1991 1991 M -> I (IN REF. 3; AAH36920).
SQ SEQUENCE 2004 AA; 230356 MW; 81124364AF791DF CRC64;

Query Match 20.8%; Score 1494; DB 1; Length 2004;
Best Local Similarity 31.7%; Pred. No. 5.5e-68;
Matches 408; Conservative 214; Mismatches 419; Indels 246; Gaps 45;

QY 78 PPLKASVENWRCVSPCLNEIDKILDCERPTKSSQSGSDARPKPIFVKQYLKWKKG 137
Db 59 FSEEDAAI-----VDKVL--SMRIVKELPSGGQYTE-----ABEFFVKYKN 97
```



```
Db 1228 RRGSGPNVKSQAQIESATRTSPAISESKKPSKDTKTETPSRQSRQSTASPVKSEK 1287
QY 1321 DDTKPOLNNVEMKDTAEETKPLR 1343
Db 1288 DGNVSLNAEQKARCKELMYPVR 1310

RESULT 12
CHD1_YEAST STANDARD; PRT: 1468 AA.
AC F32657;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE CHD1 protein.
GN CHD1 OR YER164W OR SYCP-ORF4.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313264; PubMed=919868;
RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,
RA Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
RA Hunnicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
RA Lin D., Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P.,
RA Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
RL Nature 397:78-81(1997).
CC -/- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -/- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC -/- SIMILARITY: Contains 2 chromo domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U18917; AAB64691.1; -.
DR PIR; S30818; S30818.
DR GerMOnline; 139241; -.
DR SGD; S0000966; CHD1.
DR GO; GO:0008023; C:transcription elongation factor complex; IPI.
DR GO; GO:0016887; F:ATPase activity; IDA.
DR GO; GO:0016944; F:Pol II transcription elongation factor acti...; IPI.
DR GO; GO:0006338; P:chromatin modeling; IDA.
DR GO; GO:0006368; P:RNA elongation from Pol II promoter; IPI.
DR InterPro; IPR000953; Chromo.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00385; chromo; 2.
DR Pfam; PF00271; myb DNA-binding; 1.
DR Pfam; PF00249; myb DNA-binding; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00298; CHROMO; 2.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR PROSITE; PS00598; CHROMO_1; 2.
DR PROSITE; PS00013; CHROMO_2; 2.
DR PROSITE; PS00690; DEAD ATP HELICASE; FALSE NEG.
KW Hydrolase; Helicase; Nuclear protein; ATP-binding; DNA-binding;
KW Repeat.
FT DOMAIN 195 257 CHROMO 1.
FT DOMAIN 285 350 CHROMO 2.
```

```
FT NP_BIND 401 408 ATP (POTENTIAL).
FT SITE 513 516 DEAH BOX.
SQ SEQUENCE 1468 AA; 168240 MW; 78BDB74C7FEC6BE5 CRC64;

Query Match 20.3%; Score 1461.5; DB 1; Length 1468;
Best Local Similarity 32.1%; Pred. No. 1.6e-66;
Matches 410; Conservative 217; Mismatches 403; Indels 249; Gaps 46;

QY 7 RLRIISDRKPVVNLDDSDDDFPVKPKKRTFEQVEAIVRTDAKENACQACGESTNLVSCNT 66
Db 126 RFSNRQNTVNYNDYSDDDLLLESEDD--YGESEAL-----SEENVHEA-----SANP 171
QY 67 CTYAFHAKCLV--PPLKDAVENWRCPEVCVPLNEIDKIDCEMPTKSSSEOGSSDAEPK 124
Db 172 QPEDFHGIDIVINHRLKTSLEEGKLEKTVPLDN-----NCKEN----- 210
QY 125 PIFVKQYLVKWKGLSYLHCSWVPEKEFQKAYKSNHRLK--TRVNNFHRQ-----ME 173
Db 211 ----YEFLLIKWTDESHLNTW-----ETYESIGQVRGLKRLDNYCKQFIEDQQOVRUD 259
QY 174 SFNNSE-----DFVAIR-----PEWTTVDRLACRE---EDG--ELEYLVKYKELSD 219
Db 260 PYVTAEDIEIMDMERERRLDEFEFHVPERIIDSORASLEDGTSQQLVVKWRRLNYDEA 319
QY 220 WYSESDISTFQNE-IQRFKD-VNSRTRRSKDVDRHNRNPRDFQFDHTPEFLK-GILLHPY 276
Db 320 TWENATDIVKLAPEQVKHFQRENSKILPQYSSNTSORPRPEKLSVQPPFIKGGELRDF 379
QY 277 QLEGLNFIKFSWKTQTHVLADEMGLGKTIQSIALLASLF--EENLIPLHVIAPLSTLRN 334
Db 380 QLTGTNNWAFLSKSGDNGILADEMGLGKTVQTVAFISWLI-PARRONGPHIIVVPLSTWPA 439
QY 335 WEREFEATWAPQNVVMYGTQAARAVIREHEFYLSKDQKIKKKSGQISSESKOKRIKF 394
Db 440 WLDTEKWPDLNCICYMGNKSRDTIREYEFYTNPRAK-----GKTKMF 485
QY 395 DVLTSYEMINDLSAVLPIKWECMIVDEGHRLLKNKSKLFSLSLTQYSSNHRILLTGPL 454
Db 486 NVLLTTYEYLKDRAELSGIKWQFMAVDEAHLKNAESSLYESLSFKVANRMLITGTP 545
QY 455 QNNDELFLMHFLDAGKFSLEEFQEEFKINQEQISRLHKLAPHLHRLRVKDVVKD 514
Db 546 QNNIKELAAALVNFMPGRFTIDQEIFENQDEEBEYIHDLHRRIOPTFLRLKKDKVEKS 605
QY 515 MPKKEKILRLVDLSLQKEYYKAIETRYNVQVLT--KGGAGQISLNNIMMELKVKVCHPY 572
Db 606 LPSKTERILRLVSDVQTEYYKNILTKNYSALTAKGG-HFSLNINMELKASNHPVL 664
QY 573 LEGVEPVI-----HDANFAKOLLSCGKLQLLDKMMVKLKEQGHRLVIYTOFQHML 624
Db 665 FDNASERVLPQKPGDGKMTRENVRLGLIMSSGKVVLLDQLLTRLKDGHRVLIFSQVRL 724
QY 625 DLLEDYCTHKWQYERIDKVGCGAEROIRIDFNAKNSKFCFLSLSTRAGGGINLATAD 684
Db 725 DILGDYLSIKGINFQRLDGTVPASORRISIDHFNPSDNSDFVFLSTRAGGGINLMTAD 784
QY 685 TVIIYDSWNPHADLOAMARHLGQTNKVMYRLINRGITIEERMQLTKKQVLEHLVV 744
Db 785 TVVIFSDWNFPQADLOAMARAHRIQGNHVMYRLVSKVDTVEEVLERARKKMWILEYAI 844
QY 745 G-----KLKTONIQEELDDIIRYGSKELFASEDDEAGSKGIHYDDA---AIDKLL 793
Db 845 SLGVTGDNKYTKKNEPNAGELSAILKFGAGNMTATDNQK-KLEDINLDDVLNHAEDHYT 903
QY 794 DRDLVEAEVSVDDDEENGFLKAFKANF-----EYIDENEAAALEAQRVAESKSSA 846
Db 904 TFDLGESH---LGGBE---FLKQFVTDYKADIDWDIIPEELKKLQ----- 945
QY 847 GNSDRASYWEELLKDKFELHQAEELNALGKRKRKQLV-SIEEDDLAGLEDVSSDGDS 905
Db 946 -----DEEQKRDKEYVKEQLEMMNRDNLAKKIKNSVNGDGA-----ANSDD- 991
QY 906 YEASTDGEAAGGVQVQTRRYPYRRKGRONLEPTPLMEGGRSF--RVLGFNQ----- 955
```

Db 992 ---DSTS-----RSRRRAN-DMSGSEVRALYKALFKGNLKEIDEL 1035
 QY 956 -SORAIFVOTLMRYGAGNFDWKEFVRLKQKTFEINEYGIILFKLHIAEIDENSTPT-- 1012
 Db 1036 IADGTLVPKSFKEYGE--TYDEMMEAAKDCVHEBEKN-----RKEILKLEKHATAYRA 1087
 QY 1013 ---SDGVPKLEGLRIEDVRIALLILVQKVKFVEDHPG-KPVFSTRILERFPLGRSGK- 1067
 Db 1088 KLUKGEIKAENQPKONPLRLSL--KKREKKAFLNFKGKSLNAESLLSRVEDLKLKLN 1145
 QY 1068 -----IWKEBHKIMIRAVLKHGKYGWRQAVDDKEIG 1099
 Db 1146 LINSNYKDDPLKESLGNNTPKPVQNWSSNWTKEDEKLIGVKYGYGSWTQIRDDPFLG 1205
 QY 1100 I-OLICKELNPP--HISLSAABQAGLOQNGSG--GSNPQAQTQWNPQSV----- 1145
 Db 1206 ITDKIFLNEVHPVAKKSASSDPTTPSKGKGITGS-----SKKVPGAHLGRRVDYL 1260
 QY 1146 ---ITGNNNASADGAQVNS 1161
 Db 1261 LSFLRGLGLNTKPSADIGS 1279

RESULT 13

CHD5_HUMAN
 ID CHD5_HUMAN STANDARD; PRT; 2713 AA.
 AC Q8TD26; Q8WY0; Q9H4H6; Q9H6D4; Q9NTT7; Q9P2L1;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Chromodomain-helicase-DNA-binding protein 5 (CHD-5).
 GN CHD5 OR KIAA1335.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=21886649; PubMed=11889561;
 RX Schuster E.F., Stoeger R.J.;
 RA "CHD5 defines a new subfamily of chromodomain-SWI2/SNF2-like
 RT helicases";
 RL Mamm. Genome 13:117-119(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.V., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grahnam D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Leheslahti M.H., Levensha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";

RL Nature 414:865-871(2001).
 RP [3]
 RP SEQUENCE OF 673-2713 FROM N.A.
 RX TISSUE=Brain;
 RC MEDLINE=22158633; PubMed=12168954;
 RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
 RT "Construction of expression-ready cDNA clones for KIAA genes: manual
 RT curation of 330 KIAA cDNA clones.";
 RT DNA Res. 9:99-106(2002).
 RN [4]
 RP SEQUENCE OF 1688-2713 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20181126; PubMed=10718198;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVI.
 RT The complete sequences of 150 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RT DNA Res. 7:65-73(2000).
 RN [5]
 RP SEQUENCE OF 757-1532 FROM N.A.
 RA Kawabata A., Hikiiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
 RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
 RT "NEDO human cDNA sequencing project.";
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE OF 2440-2713 FROM N.A.
 RC TISSUE=Lymph;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Marra M.A.;
 RA Schnerch A., Schein J.E., Jones S.J.M., Skalska U., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: Probable transcription regulator.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- TISSUE SPECIFICITY: Ubiquitous.
 CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
 CC -!- SIMILARITY: Contains 2 chromo domains.
 CC -!- CAUTION: Ref.5 sequence differs from position 1526 onward for
 CC unknown reason.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AY034072; AAK56405.1; -
 CC EMBL; AL031667; -; NOT ANNOTATED_CDS.
 CC EMBL; AL031669; -; NOT ANNOTATED_CDS.
 CC EMBL; AL121674; BAA90808.1; -
 CC EMBL; AB037756; BAA92573.2; -
 CC EMBL; AK026022; BAB15325.1; ALT_SEQ.
 CC EMBL; BC021907; AAB12907.1; -
 CC Genew; HGNC:16816; CHD5.

DR GO: GO:0005634; C:nucleus; NAS.
 DR GO: GO:0003682; F:chromatin binding; NAS.
 DR GO: GO:0006338; P:chromatin modeling; NAS.
 DR GO: GO:0007399; P:neurogenesis; NAS.
 DR GO: GO:0006355; P:regulation of transcription, DNA dependent; NAS.
 DR InterPro: IPR000953; Chromo.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002464; DEAD box.
 DR InterPro: IPR001650; Helicase C.
 DR InterPro: IPR001005; Myb DNA Binding.
 DR InterPro: IPR000330; SNF2_N_Binding.
 DR Pfam: PF00271; Helicase C; 1.
 DR SMART: SM00176; SNF2 N; 1.
 DR SMART: SM00298; CHROMO; 2.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 DR SMART: SM00717; SANT; 1.
 DR PROSITE: PS00598; CHROMO_1; FALSE NEG.
 DR PROSITE: PS00613; CHROMO_2; 1.
 DR PROSITE: PS00690; DEAD ATP HELICASE; 1.
 KW Chromatin regulator; DNA-binding; ATP-binding; Helicase;
 KW Nuclear protein; Repeat; Transcription regulation.
 FT DOMAIN 290 341 CHROMO 1.
 FT NP BIND 373 437 CHROMO 2.
 FT SITE 596 599 DEAD BOX.
 FT DOMAIN 84 89 POLY-GLY.
 FT DOMAIN 91 200 LYS-RICH.
 FT DOMAIN 1864 1874 POLY-GLU.
 FT DOMAIN 2283 2289 POLY-ARG.
 FT DOMAIN 2652 2659 POLY-LYS.
 FT CONFLICT 132 132 E -> K (IN REF. 2; CAB90808).
 FT CONFLICT 1029 1029 K -> E (IN REF. 3 AND 5).
 FT CONFLICT 2440 2445 RGRPR -> EIVGLE (IN REF. 6).
 FT CONFLICT 2661 2661 D -> G (IN REF. 1).
 SQ SEQUENCE 2713 AA; 305149 MW; FFA07E22E01DAAAF CRC64;

Query Match 19.8%; Score 1422.5; DB 1; Length 2713;
 Best Local Similarity 31.2%; Pred. No. 3.5e-64;
 Matches 391; Conservative 216; Mismatches 425; Indels 223; Gaps 44;

QY 14 RKPVNLDSDDDFPVKDRTPEQVE-----AIVRTAKENACQAGESTNLVS 63
 DB 220 RSPEESTSTDS-----QKRSGQVRKRYNEDLDFKVVDDGDTIAVLGAGRTSAL-- 272

QY 64 CNTCTVAFHAKLVPLPKDASVENWCPCVSPLEIDKTLDCMEPTKSSQSGSDAP 123
 DB 273 -SASTLAWQAE--EPEDDANI-----IEKIL-----ASKTVQEVHPGEP 309

QY 124 KPIFVQYLVKGLSYLHCSWVPEKEFOKAYKSNHRLKTRVNNFHRQMES-FNNSEDDF 182
 DB 310 -PFDLELFYVYKRNFSYLCHKWATMELEKDPRIQKIK-RPNKQAKMKHIFTEPDEDL 367

QY 183 VAIRPENTVDRIACRE-----EDGE--LEYLVKYSYDECYWESESDISTFQNEIOR 236
 DB 368 --FNPDYVEVDRIEVAHTKDAETGBEVTHLVKWCSLPYEESTWLEEDVD--PAKVKE 423

QY 237 FKVNSNTRRSKVDHKKRPR--DFQCFDHTPEFL-KGLLHPYQLEGLNLFPSWSKQTH 293
 DB 424 FESL----QVLPIKHYVERPASWSQKLEKRSYKNSQLREYQLEGMNWLFPNNYRNKN 479

QY 294 VILADEMGLKTTQSIALLASLPEENLI-PHLVIAPLSTLRNWEREFATWAPQNVVMYF 352
 DB 480 CILADEMGLKTTQSIITFLSEIFLRGHPFLIAPLSTITNWEREFRTWT-EMNAIVYH 538

QY 353 GTQAARAVIREHFYLSKQKTKKKKSGQISSEKQKRIKPDVLTSYEMINLDSAVLK 412
 DB 539 GSQISRMQIQYEM-----VYRDAQGNPLSGV-----FKFHWITTFEMILADCPCLK 586

QY 413 PIKWCQIVDEGHLKKNKSKLFSLLTQVYSSNRIILLTGTPLONNDELFLMHHFLDACK 472
 DB 587 KIHWSCVILDEAHLRKNRNKCLLEGLKIMALEHKVILLTGTPLONSVEELFSLNLFLEPSQ 646

QY 473 FGSLEEFQEEFKDINQERQISRLHQMGLAPHLRLRRVKDVMKMDPPKELLILRLVDSLSQK 532
 DB 647 FPSETAFLEEFGLKTERQVKVQLQSILKPMMLRLKDDVEKNLAPKQETIIEVELTIQK 706

QY 533 EYKKAIFRNYQVLTCKGGAQISLNNI---NMELRKVCVCHPVMYLGCV-EPVHIDANEAA-- 586
 DB 707 KYRAILEKNFNSFLT-KGANQHNMPLNLTMMELRKCCHHPYLINGAEKILEDRKTKTHS 765

QY 587 -----FKOLLESCGKLQLLDKMMVKLKEQCHRVLYITQFOHMLDLDLEDYCTHKKWQYE 639
 DB 766 PDAPDFQIQAMIQAAAGKLVLDKLLPKLIAGGHKVLIFSQWVRCILDILEDYLIQRYIYE 825

QY 640 RIDGKVGGAERQIRIDRFNAKNSKFCFLSTRAGGLGINLATATVYIYDSWNPHADL 699
 DB 826 RIDGRVGNLRQAAIDRFCKPDSRDFVFLCTRAGGLGINLTAATCIIIFSDMNPQNDL 885

QY 700 QAMABAHRLGQTNKVMYIYRLNRTIERNMQLTKKKWVLEHLV-----GLKLTQNI 752
 DB 886 QAQARCHRIGOSKAVKYVRLITRNSYEREMFDKSLKGLGDKAVLQDINRKGTTNGVQOL 945

QY 753 NOEELDDIIRYGSKELFASEDEAGSKGIHYDDAAIDKLDRLDVEAEVSVDDDEENG 812
 DB 946 SKMEVEDLLRKAGYCALMDEDEGSK-----FCEEDIDQILQR---RHTTITIQSE---G 994

QY 813 FLKAFKVFANFEYIDENAAALEAQVAAESKSSAGNSDRASWYELLK-DXPELHQABEL 871
 DB 995 KGSTFPAKASF-----VASGNRTDI-SLDDPNPQWKWAKIADLDEAKNEK 1038

QY 872 NALG-KRKRSRKQLV---SIEEDDLALEDVSSDGDSEYAEESTDGEAAGQGVQGRBPY 927
 DB 1039 ESLVIDRPRVRKQTKHYNSEFEDELMEFSELOSDSDER-----PTRRRINDKARRY 1090

QY 928 RKGRDNLNPTPLMEGEGSRFVLGFGNSQ-----RAIFVQTLRMRYGAG---- 971
 DB 1091 LBAECFRVEKNLLIFGWRGKDLITGFRKWLNEKDMEMICRALLVYCVKHYKDEKIK 1150

QY 972 NPDWKEFVP-----RLKQKTFEBI-----NEVGI 995
 DB 1151 SPIWELITPTKDGQAQTLQNHSGLSAPVPRGRKGGKTKNQLLIPELKDADMLATCNPEV 1210

QY 996 L-----FLKHIABEIDENSPTFSDGVPKEGLRLEDVLRIALLLVQEKVKFVEDHPGFPV 1051
 DB 1211 LHDDGYKKHLKQHCNK-----VLLKVRMLDYLUKAEILGBAAEKAFEG 1252

QY 1052 FPSRILE-RFPGLRSGKI---WKEEHDKIMIRAVLKHGVRQWQAIYVDDKELIGELICK 1106
 DB 1253 SPARELDVLPDIDYWEIPVDWMDAEADKSLIIGVFKHYERYNAMRADPALCFLE---- 1308

QY 1107 ELNFPHISLSAAEQAGLQ-----QNGSGGSPGAGTQNPQSGSVITGNNNASADG 1156
 DB 1309 KVGMPDEKLSAEQGVTDGTSIDIPERGNTDKEDNAEDKVDGIQKOTESSSSDGGDG 1363

RESULT 14
 HRP1_SCHPO
 ID HRP1_SCHPO STANDARD; PRT; 1373 AA.
 AC Q9US25; Q92369; Q9U28;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Chromodomain helicase hrp1.
 GN HRP1 OR SPAC1783.05.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 ON NCBI_TaxID=4896;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JY741;
 RX MEDLINE=98179088; PubMed=9520266;
 RA Jin Y.H., Yoo E.J., Jang Y.K., Kim S.H., Kim M.J., Shim Y.S.,
 Lee J.S., Choi I.S., Seong R.H., Hong S.H., Park S.D.;

578 RDELNFQDPAEQERDIRDQLQRPFLRLRLKDKVEKSLPSKSEIRILVELSDMQTWEY 637
 536 KALFTNYQVLT--KGGQAIISLNNIMELRKVCCHPYMLEGVPV-----IHDNEAFK 588
 638 KNILTKNYRALTGHTDGRGQLSLNLVVELKRVSNHPYLPFGAAEKMMGRKWTREDTLR 697
 589 QLLESCGQLQLDKMMVKLKEQGRVLIYTFQFMDLLELDYCTHKKWOYERIDGRVGA 648
 698 GIIMNSGKVLKLLQRLKHGHRVLIIFSQVYRMLNIGELYSRGNVYQRLDGTIPAS 757
 649 ERQIRIDRNKNSKFCFLSLSTRAGGLGINATADTVIYSDMNPADLQAMARHL 708
 758 VRRVSIHDHFNAPSPDFVLLSTRAGGLGINLTADTVIIFSDMNPQADLQAMARHRI 817
 709 GQTNKMWYRLNIRGFIIEERMMQLTKKMWLHLVVG-----KLQTNQNEELDI 760
 818 GQKNHNVYRFLSKDTEEDILERRARRKMILEYIISLVGTEKSKNSKNDKYDAQLSAI 877
 761 IRYGSKELFASDEDEAGKGIHYDDAAIDKLDRDLVEAEV---SVDDSEBENGFLKAF 817
 878 LKFGASWPKATENQK-KLENMNLDD-ILSHAEDRD--SSNDVGGASMGEE---FLKQF 930
 818 KVAFYVIDENEAALAEQRAVAESKSSAGNSDRASYWELLLKDPHQAELNALGKR 877
 931 EVTDYKAEDLN-----WDDIIPEE-EMERIEEERMLAA 963
 878 KRSRQLVSIIEEDDLAGLEDVSSDGSYEAESTDGEAAGQGVQTCRRPYRKRGRDNLFP 937
 964 QRAK-----EERERREERENDEHDPS-----RTYKRTKS-----996
 938 TPLMEGEGRSFVLGNQSORAIFVQTLMYRGAGNFDWKEFVPLRKQKTFEINEYGIIL- 996
 997 --ITKQORRENV--REKEIRLLYRAMIKFGLVDERPDTIVKEAELOATDPKRIYLSLA 1052
 997 -FLKHAIEIDE-NSPTFSQVPEKGLRI-----EDLVRLTALLILVOEKVKFVE 1044
 1053 DMVKACDEAVERLGADDTKNQPKAILIEFKGVKNINAEVTVLRVKDLTHLHRAVYKGLD 1112
 1045 DHPGKPV--PPSILRFPCLRSKTKKEHDKIMIRAVLKHGCGYRQWQIVDDKEIGIOE 1102
 1113 --FLKQIIGPITRSHVW---NCSWGIKEDSMLLAGINKHGFQWQIAKNPDJLGLHD 1165
 1103 LIC-----KELNF-----PHISLSAAEQAGLQGGQSGSNP 1134
 1166 KIFLDEAKNDKESRYPVSAHLVRRGEYLLSVVREHPDLFVVKTDQ-----P 1212
 1135 GAGTQNPQSVITGNNNASADGAQVNSMFYRRQMRRLVEFVKRVLLEKAMNVEYABE 1194
 1213 TKRYNRKAPTKSSTRQTTLDGSIKNTKSSRTKXK-----BE 1251
 1195 YVGLGSSSTPTEPEAEKPIADTVGSFTEVDDEMLDGLPKTDPTITSEIMGAAVDNNQ 1254
 1252 ETVRGDETS-----PEG-----TVGED--EVEEPRQAEPPKRALRSN--GKAASNR 1296
 1255 ARVEIAQHYNQMKLLDENARESVQAYVNNQPTSTKVNESFRALKSGING-NINTILSITS 1313
 1297 TTRNSMKTHSAMDTL---TAAVALDAELDNW-SNEKAKEVDHVKSENGESVNE--PNT 1350
 1314 DQSKSHEDDTKPOLNINVMKDTAEETKPL 1342
 1351 DLSLETEENT-----TVSDISPL 1368

RESULT 15

SN21_HUMAN STANDARD; PRT; 976 AA.
 ID SN21_HUMAN AC P28370;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Possible global transcription activator SNF2L1.
 GN SMARCA1 OR SNF2L1 OR SNF2L.
 OS Homo sapiens (human).

Query Match 15.8%; Score 1139; DB 1; Length 976;
 Best Local Similarity 32.2%; Pred. No. 2.1e-50;
 Matches 301; Conservative 160; Mismatches 311; Indels 164; Gaps 24;

QY 223 SESDSTFQNEIQRFKDVNSRTRRSKDVHKKENPRDFQFQDHTPELK-GLLHPYQLEGL 281
 DB 62 SAGDYRHRTEQEEDEELSEKTSNV-----CINFEVSPSVKGGPLRDYQIRGL 113
 QY 282 NPLRFSWSQTHVILLADEMGLKTIQSIALLASLFEENLI---PHLVIAPLSLTLRNWEREF 339
 DB 114 NMLISLYENGVLGILADEMGLKTIQTALLGLYKHYRNIPGFHVLVFKSLHNMWNEF 173
 QY 340 ATWAPQMVVYFGTAQAPAVIREHEFYLSKDOKKIKKKKSGQISSESKQRIKFDVLLT 399

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN NCBI_TaxID=9606;
 RP [1]
 RC SEQUENCE FROM N.A.
 RX TISSUE=Kidney;
 RA MEDLINE=93027164; PubMed=1408766;
 RA Okabe I., Bailey L.C., Attree O.F., Perkel J.M., Nelson D.L.,
 RA Nusbaum R.L.;
 RT "Cloning of human and bovine homologs of SNF2/SWI2: a global
 RT activator of transcription in yeast *S. cerevisiae*.";
 RL Nucleic Acids Res. 20:4649-4655(1992).
 RN [2]
 RP PHOSPHORYLATION OF TYR-876.
 RX MEDLINE=22107313; PubMed=12112843;
 RA Maguire P.B., Wynne K.J., Harney D.F., O'Donoghue N.M., Stephens G.,
 RA Fitzgerald D.J.;
 RT "Identification of the phosphotyrosine proteome from thrombin
 RT activated platelets.";
 RL Proteomics 2:642-648(2002).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family. SNF2L
 CC subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; M88163; AAA80559.1; -
 DR EMBL; M89907; AAA80560.1; ALT_INIT.
 DR FIR; S35457; S35457.
 DR Genew; HGNC:11097; SWARCA1.
 DR MIM; 300012; -
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0004386; F:helicase activity; TAS.
 DR GO; GO:0006338; P:chromatin modeling; TAS.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR001005; Myb_DNA_Binding.
 DR InterPro; IPR00330; SNF2_N.
 DR Pfam; PF00271; Helicase_C_1.
 DR Pfam; PF00249; myb_DNA-Binding; 2.
 DR Pfam; PF00176; SNF2_N; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELICC; 1.
 DR SMART; SM00717; SANT; 2.
 DR PROSITE; PS00690; DEAD_ATP_HELICASE; FALSE NEG.
 KW Transcription regulation; Hydrolase; Helicase; Activator;
 KW Nuclear protein; ATP-binding; Phosphorylation; Polymorphism.
 FT NP_BIND 130 137 ATP (POTENTIAL).
 FT SITE 233 236 DEAD BOX.
 FT MOD_RES 876 876 PHOSPHORYLATION.
 FT VARIANT 578 578 R -> Q.
 FT /FTID=VAR_001242.
 SQ SEQUENCE 976 AA; 114732 MW; CC91633E24341B64 CRC64;

Db 174 KRWPSLRVICFGDKDARAARFIRDEMMPG-----EWDVCVT 210
Qy
400 SYEMINLDSAVLKPIKWEKMWVDEGHRLLKNKDSKLFSSLTQYSSNHRILLTGTPLQNNLD 459
Db 211 SYEWIKESKVPKFWYRLVDEAHRIRKNEKSKLSEIVREFKSTNRLLLTGTPLQNNLH 270
Qy 460 ELFPLMLHFLDAGKFGSLEEFQEFKDIN--QEEQISRLHKMLAPHLARRVKKDVWKDMP 516
Db 271 ELWALLNFLPDPVFNSSADDFDSWFDTKNCLGQOKLVERLHVLKPLRRIRKTDVEKSLP 330
Qy 517 PKKELILRVDLSLQKEYYKAIFTRNQVLTQKKGGA-QISLNNIMMELRKVCCHPYMLEG 575
Db 331 PKKEIKYILGSLQWQEWTKILMKDIDVLNSSGKMDKRWLLNLLMQLKCCNHPYLPFDG 390
Qy 576 VEPVIHDANEAFKQLLESCKQLQLDKMMVKLKEQGHVRVLIYTOQHMLDLLEDYCTHKK 635
Db 391 ABPGPPYTTD--EHIVNSGKMWLDKLAKLKEQGSRVLFESQMTRLDILEDCMWRG 448
Qy 636 WYERIDGKVGGAERQIR-----IDRENKSNKECFLLSTRAGGLGINLATA 683
Db 449 YEYCRLDGQTPHEEREDKLEVEFEGQREAEAFNAPNSKFFIMLSTRAGGLGINLASA 508
Qy 684 DTVIYDSDWNPADLOAMARHRLQOTNKVMYRLINRGTIEERMWOLTKKMWLEHLV 743
Db 509 DVVILYDSDWNPQVDLOAMDRAHRIGQKPVVFRLLITNTVEERIVERAEIKRLDSIV 568
Qy 744 V--GKLTQTNIN--QEELEDDIIRYGSKELFASDEDEAGSKGKIHYDDAAIDKLDRLDV 798
Db 569 IQQRLIDQRSNKLAKEMLMQIRGHATHVFASKES-----LTDEDTITILERGEK 620
Qy 799 EAEVVS-----VDDEENGFLKAFKVFANFEYIDENEAALAEQORVAESK 843
Db 621 KTAEMNERLQKMGESSLRNFRMDIEQ-----SLYKFEGEDYREKQKLGWVE--WIEPPKR 673
Qy 844 SSAGNSDRASYWBEELLKDK-----FELHOAEEL--NAL 874
Db 674 ERKANYAVDAYFREALRVSEPKIPKAPRPPKQPNVDQFQFPFRLFELLEKEILYRKYI 733
Qy 875 G-KRKRSRKQVLSIBEDDLAGLEDVSS-DGDESIEAESTDGEAAGQVGTGRRPYRRKGR 932
Db 734 GYKVPN-----PDIPNALAQREEQKIDGAEPLTPEETE----- 769
Qy 933 DNLEPTPLMEGSRSPRVLGFNQSQRAIFVQTLMYGAGNPDWKEFVPRLKOKTFEEINE 992
Db 770 ----EKEKLTQGTNTWTKRDFNQ-----FIKANEKYGRDDID--NIAREVEGKSPEEYME 819
Qy 993 YGILFLKHTABEIDENSPTFSDGVPKEGLRIEDVLVRIALLILOEKVKFVEDHFGKPVF 1052
Db 820 YSAVFWERCNELOD-----IEKIMAQIERGEARIORRISIK---KALDAKIARYKA 867
Qy 1053 PSRILRFPPLRSKGIWKEEHDKIMIRAVLKHGYGR 1088
Db 868 PFHQLRIQYGTSGKNYTEEDRFLICMLHKMGFDR 903

Search completed: September 14, 2004, 02:04:58
Job time : 45 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein. - protein search, using sw model
Run on: September 14, 2004, 01:52:16 ; Search time 162 Seconds
(without alignments)
2695.538 Million cell updates/sec

Title: US-10-049-137-2
Perfect score: 7187
Sequence: 1 MSSLVERLIRSDRKPYNL.....VDVMEAEKBEKPKNMVVD 1384

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phase.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteria.*
 - 17: sp_archaea.*

SUMMARIES					Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.	
Result No.	Score	Query Match	Length	ID	Description	
1	7187	100.0	1384	10	Q9S775	Q9S775 arabidopsis
2	3114	43.3	1067	10	Q9SZ57	Q9sz57 arabidopsis
3	2105.5	29.3	1954	4	Q8TD10	Q8td10 homo sapien
4	2080	28.9	1937	4	Q81XZ5	Q8ixz5 homo sapien
5	2076	28.9	1893	13	Q7ZWN3	Q7zwn3 xenopus lae
6	2066	28.7	2000	4	Q9Y410	Q9y410 homo sapien
7	2048.5	28.5	1829	5	Q19815	Q19815 caenorhabdi
8	1934.5	26.9	1388	4	Q9UFR9	Q9ufr9 homo sapien
9	1805	25.1	1045	11	Q8BM83	Q8bm83 mus musculu
10	1578	22.0	1808	13	Q82142	Q82142 gallus gall
11	1572	21.9	2259	10	Q7XIF9	Q7xif9 oryza sativ
12	1565.5	21.8	1738	10	Q9SI41	Q9si41 arabidopsis
13	1554	21.6	5322	5	Q9VPL9	Q9vpl9 drosophila
14	1553.5	21.6	1786	13	Q7ZT34	Q7zt34 poephila gu
15	1551	21.6	1806	13	Q800S6	Q800s6 poephila gu
16	1537	21.4	1461	5	O17909	O17909 caenorhabdi

17	1522.5	21.2	1883	5	Q24376	Q24376 drosophila
18	1505	20.9	2228	10	Q48579	Q48579 arabidopsis
19	1496.5	20.8	1081	11	Q8C7Q2	Q8c7q2 mus musculu
20	1490.5	20.7	981	13	Q9DQ02	Q9dqq2 gallus gall
21	1456	20.3	918	13	Q9DQ03	Q9dgg3 aegolius fu
22	1437	20.2	918	13	Q9DQ04	Q9dgg1 nymphicus h
23	1437	20.0	918	13	Q9DQ04	Q9dgg4 aegolius fu
24	1434	20.0	918	13	Q9DQ00	Q9dgg0 nymphicus h
25	1345	18.7	1101	5	Q8SWV9	Q8swv9 drosophila
26	1332.5	18.5	2957	5	Q61845	Q61845 caenorhabdi
27	1290	17.9	1141	10	Q8LJ77	Q8lj77 oryza sativ
28	1263.5	17.6	1251	5	Q8SWP7	Q8swp7 encephalito
29	1258	17.5	3328	5	Q81JG6	Q81jg6 plasmodium
30	1224.5	17.0	1057	10	Q9SQ03	Q9sq03 arabidopsis
31	1220.5	17.0	1055	10	Q8RWY3	Q8rwy3 arabidopsis
32	1219.5	17.0	1137	10	Q9FRB8	Q9frb8 oryza sativ
33	1215	16.9	1122	10	Q93VU7	Q93vu7 oryza sativ
34	1195.5	16.6	891	11	Q8BI88	Q8bie8 mus musculu
35	1173.5	16.3	1418	3	Q9P793	Q9p793 schizosacch
36	1157	16.1	730	4	Q9H9V7	Q9h9v7 homo sapien
37	1151.5	16.0	1052	4	Q60264	Q60264 homo sapien
38	1146.5	16.0	1051	11	Q91ZW3	Q91zw3 mus musculu
39	1143.5	15.9	1064	11	Q91Y58	Q91y58 mus musculu
40	1141.5	15.9	1051	11	Q925M9	Q925m9 mus musculu
41	1141	15.9	1102	5	Q7YU02	Q7yyu2 cryptospori
42	1139	15.8	1120	3	Q08773	Q08773 saccharomyc
43	1136	15.8	1046	11	Q8BSS1	Q8bsel mus musculu
44	1136	15.8	1046	13	Q9DF71	Q9df71 xenopus lae
45	1132	15.8	1027	5	Q8SX14	Q8sx14 drosophila

ALIGNMENTS

RESULT 1	
Q9S775	
ID	Q9S775 PRELIMINARY; PRT; 1384 AA.
AC	Q9S775;
DT	01-MAY-2000 (Tremblrel. 13, Created)
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT	01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE	Chromatin remodeling factor CHD3.
GN	PKL
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.
OX	NCBI_TaxID=3702;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=cv. Columbia;
RX	MEDLINE=20040637; PubMed=10570159;
RA	Ogas J., Kaufmann S., Henderson J., Somerville C.;
RT	"PICKLE is a CHD3 chromatin-remodeling factor that regulates the
RT	transition from embryonic to vegetative development in Arabidopsis."
RL	Proc. Natl. Acad. Sci. U.S.A. 96:13839-13844(1999).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=cv. COLOMBIA;
RA	Eshed Y., Baum S.F., Bowman J.L.;
RT	"Distinct mechanisms promote polarity establishment in carpels of
RT	Arabidopsis thaliana."
RL	Cell 0:0-0(1999).
DR	EMBL; AF185577; AAF13875.1; -.
DR	EMBL; AF185578; AAF07084.1; -.
DR	PIR; T52301; T52301
DR	GO; GO:0000785; C:chromatin; IEA.
DR	GO; GO:0005634; C:nucleus; IEA.
DR	GO; GO:0005524; F:ATP binding; IEA.
DR	GO; GO:0008056; F:ATP dependent helicase activity; IEA.
DR	GO; GO:0003682; F:chromatin binding; IEA.
DR	GO; GO:0016787; F:hydrolase activity; IEA.
DR	GO; GO:0006333; F:chromatin assembly/disassembly; IEA.

GO: 0006355; P: regulation of transcription, DNA-dependent; IEA.
 DR InterPro: IPR000953; Chromo.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR003330; SNF2_N.
 DR InterPro: IPR001965; Znf_PHD.
 DR InterPro: IPR001841; Znf_ring.
 DR Pfam: PF00385; chromo; 2.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00628; PHD; 1.
 DR Pfam: PF00176; SNF2_N; 1.
 DR SMART: SM00298; CHROMO; 2.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 DR SMART: SM00249; PHD; 1.
 DR SMART: SM00184; RING; 1.
 DR PROSITE: PS00013; CHROMO_2; 2.
 DR PROSITE: PS01359; ZF_PHD_1; 1.
 DR PROSITE: PS00016; ZF_PHD_2; 1.
 KW ATP-binding; Helicase; Hydrolase.
 SQ SEQUENCE 1384 AA; 158404 MW; C4EDC75D7D973264 CRC64;

Query Match 100.0%; Score 7187; DB 10; Length 1384;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1384; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLSVERLRIRSDRKPVNLDSDDDDFVPKDRTEQVEAIVRTDAKENACQAGESTN 60
 DB 1 MSLSVERLRIRSDRKPVNLDSDDDDFVPKDRTEQVEAIVRTDAKENACQAGESTN 60

QY 61 LVSNCNTCTYAFHAKCLVPLPKDASVENWRCPCVSPLEIDKILDCEMRPTKSSEQSSD 120
 DB 61 LVSNCNTCTYAFHAKCLVPLPKDASVENWRCPCVSPLEIDKILDCEMRPTKSSEQSSD 120

QY 121 ABPKPIFVKQYLVKWKGLSYLHCSWVPEKEFOKAYKSNHRLKTRVNNPHROMESFNNSD 180
 DB 121 ABPKPIFVKQYLVKWKGLSYLHCSWVPEKEFOKAYKSNHRLKTRVNNPHROMESFNNSD 180

QY 181 DFVAIRPEWTVDRILAIACREDEGELEYLVKYKELSYDECWSESDISTFQNEIORFKDV 240
 DB 181 DFVAIRPEWTVDRILAIACREDEGELEYLVKYKELSYDECWSESDISTFQNEIORFKDV 240

QY 241 NSRTRSKVDHKNRPDRFQQDHTPEFLKGLLHPYQLEGLNPLRFPSWSKQTHVILADEM 300
 DB 241 NSRTRSKVDHKNRPDRFQQDHTPEFLKGLLHPYQLEGLNPLRFPSWSKQTHVILADEM 300

QY 301 GLGKTTQSIALLASLFEENLIPLHVIAPLSTLRNWEREFATWAPQNVVYFGTAQRAV 360
 DB 301 GLGKTTQSIALLASLFEENLIPLHVIAPLSTLRNWEREFATWAPQNVVYFGTAQRAV 360

QY 361 IREHEFYLSKQOKKIKKKSGQISSESOKRIKFDVLLTSYEMINLDSAVLKPKEWCM 420
 DB 361 IREHEFYLSKQOKKIKKKSGQISSESOKRIKFDVLLTSYEMINLDSAVLKPKEWCM 420

QY 421 VDEGHLKKNKDSKLFSSLTQYSSNHRILLTGTPLNQNLDELFLMHFLDAGFGSLEEFQ 480
 DB 421 VDEGHLKKNKDSKLFSSLTQYSSNHRILLTGTPLNQNLDELFLMHFLDAGFGSLEEFQ 480

QY 481 EEPKDNQBEQISRLHKLAPHLRLRVKQDVKMDMPKKEILIRVDLSSLOKEYYKAIPT 540
 DB 481 EEPKDNQBEQISRLHKLAPHLRLRVKQDVKMDMPKKEILIRVDLSSLOKEYYKAIPT 540

QY 541 RNYQVLTQKGGAGQISLNNIMELRVKVCCHPYMLEGVEPVTHDANEAFKQLLESCKQL 600
 DB 541 RNYQVLTQKGGAGQISLNNIMELRVKVCCHPYMLEGVEPVTHDANEAFKQLLESCKQL 600

QY 601 DKMWVKLEQGHRLVYTFQHMLEDLEDYCTHKKWQYERIDGKVGGAERQIRIDRFNAK 660
 DB 601 DKMWVKLEQGHRLVYTFQHMLEDLEDYCTHKKWQYERIDGKVGGAERQIRIDRFNAK 660

QY 661 NSNKFCLSTRAGGLGINLATADTVIYDSDWNPHADIQAMARAHLQGTNKVMYIRLI 720
 DB 661 NSNKFCLSTRAGGLGINLATADTVIYDSDWNPHADIQAMARAHLQGTNKVMYIRLI 720

QY 721 NRGTTIEERMVQLTKKKVLEHLVVGKLTQINQBELDDIIRYGSKELFASEDDAGKSG 780
 DB 721 NRGTTIEERMVQLTKKKVLEHLVVGKLTQINQBELDDIIRYGSKELFASEDDAGKSG 780

QY 781 KIHYYDAAIDKLLDRDLVEABEVSVDDEENGFLKAFKVANPEYIDENEAALAEQRAVA 840
 DB 781 KIHYYDAAIDKLLDRDLVEABEVSVDDEENGFLKAFKVANPEYIDENEAALAEQRAVA 840

QY 841 ESKSSAGNSDRASYWEELLKDFELHQAEEALNALKRKRKQLVSIIEEDDLAGLEDVSS 900
 DB 841 ESKSSAGNSDRASYWEELLKDFELHQAEEALNALKRKRKQLVSIIEEDDLAGLEDVSS 900

QY 901 DGDSEYEAESTDGEAAGQVOTGRPYRRKGRDNLEPTPLMEGEGRSFRVLGFGNSQRAI 960
 DB 901 DGDSEYEAESTDGEAAGQVOTGRPYRRKGRDNLEPTPLMEGEGRSFRVLGFGNSQRAI 960

QY 961 FVQTLMYRGAGNFDWKEFVPRLKQKTFBEINEYIGILFLKHAIEIDEINSPTFSDGVPK 1020
 DB 961 FVQTLMYRGAGNFDWKEFVPRLKQKTFBEINEYIGILFLKHAIEIDEINSPTFSDGVPK 1020

QY 1021 LRIEDVLVRIALLILVQEKVFEVDHPGKVPFPPSILRFPGLRSKGIWKEHDKIMIRA 1080
 DB 1021 LRIEDVLVRIALLILVQEKVFEVDHPGKVPFPPSILRFPGLRSKGIWKEHDKIMIRA 1080

QY 1081 VLKHGCGRWQAIIVDDKELGIGELICKELNFPHISLSAAEQAGLQCGQSGSNPGAQTNQ 1140
 DB 1081 VLKHGCGRWQAIIVDDKELGIGELICKELNFPHISLSAAEQAGLQCGQSGSNPGAQTNQ 1140

QY 1141 NPGSVITGNNASADGAQVNSFFYRDQMRRLVFEVKKRVLLLEKAMNYEYAEYYGLGG 1200
 DB 1141 NPGSVITGNNASADGAQVNSFFYRDQMRRLVFEVKKRVLLLEKAMNYEYAEYYGLGG 1200

QY 1201 SSSIPEEPEAPKTIADTVGVSFIEVDDMLDGLPKTDPITSEETMGAAVDNNQARVEA 1260
 DB 1201 SSSIPEEPEAPKTIADTVGVSFIEVDDMLDGLPKTDPITSEETMGAAVDNNQARVEA 1260

QY 1261 QHYNQMKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNINTILSIITSDQSKSHE 1320
 DB 1261 QHYNQMKLLDENARESVOAYVNNQPPSTKVNESFRALKSINGNINTILSIITSDQSKSHE 1320

QY 1321 DDTKPDNLNNVEMKDTAETKPLRGVVDLNVVGEENIAEASGVVDVVKMEAEKBEKPKN 1380
 DB 1321 DDTKPDNLNNVEMKDTAETKPLRGVVDLNVVGEENIAEASGVVDVVKMEAEKBEKPKN 1380

QY 1381 MVVD 1384
 DB 1381 MVVD 1384

RESULT 2

Q9SZ57 PRELIMINARY; PRT; 1067 AA.
 ID Q9SZ57;
 AC Q9SZ57;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein
 GN Flic18.100 OR AT4G31900.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosoids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RP SEQUENCE FROM N.A.
 RA Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,
 RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
 RA Villarroel R., Gielen J., Van Montagu M., Hoheisel J., Mewes H.W.,
 RA Mayer K.F.X., Schueller C.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.


```

[1]
RP SEQUENCE FROM N.A.
RA Thompson P.M., Gotch T., White P.S., Brodeur G.M.;
RT "CHD5, a New Member of the Chromodomain Gene Family, is Preferentially
RT Expressed in the Nervous System.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF425231; AAL98962.1;
DR GO; GO:000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0003682; F:chromatin binding; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0006333; P:chromatin assembly/disassembly; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000953; Chromo.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002464; DEAH box.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2 N.
DR InterPro; IPR001965; Znf PHD.
DR Pfam; PF00385; helicase_C; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00628; PHD; 2.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00298; CHROMO; 2.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS00013; CHROMO_2; 2.
DR PROSITE; PS00690; DEAH ATP HELICASE; 1.
DR PROSITE; PS00016; ZF PHD 2; 2.
KW ATP-binding; Helicase; Hydrolase.
SQ
SEQUENCE 1954 AA; 223048 MW; E333062B5B5E71F CRC64;

Query Match      29.3%; Score 2105.5; DB 4; Length 1954;
Best Local Similarity 35.9%; Pred. No. 2.7e-109;
Matches 533; Conservative 227; Mismatches 437; Indels 289; Gaps 49;

QY 21 DSDSDDDFPVKDRTEQVEAIVRTDAKENAQAQGESNLVSCNTCTVAFHAKLVLPPL 80
DB 395 EPKDDDD--EEBEGGCEEE-----DDHMFRCVCKGGELICCDACPSYHLHCLNPPL 447

QY 81 KDAVENWRCPECVSP--LNEIDKILDCE-MRPTKSSEQ--SSDAE-----PKP---IF 127
DB 448 PEIPNGEWLCPCTCPPLKGKQVRIHWEWTEPPAPFMVGLPGDPVPSLPPLPKLEGIP 507

QY 128 VKQYLVKWKGLSYLHCSWVPEKEFQKAYKSNHRLKTRVNNFHR-----QMES 174
DB 508 EREFVVKWAGLSYHCSWVKELQLELYHTVMYRNQKNDMDPEPPFDYSGDGDGKSK 567

QY 175 FNNSS-----EDDF--VAIREWTVDRIL-ACREEDGELEVLVYKELSYDECYWE-S 223
DB 568 RKNKDPFLYAKMEERFYGIKPEWMIHRLNHSFDKGDGVHYLIKWKDLPYDQCTWEID 627

QY 224 ESDISTFQNEIORF-----KDVNSRTRRSKDVHDKNRPDRFQOFDHTPEF 268
DB 628 DIDIPYDNLKQAYWGHRELMLGEDTRLPKLLKGGKLRD-DKQEKPPDTPIDVPTVKF 686

QY 269 LK-----GLLHPYQLEGLNFRFSWSKQTHVLADEMGLGKGTIQSIALLASLFE- 318
DB 687 DKQPMWIDSTGGTLPYQLEGLNFRFSWAQGTDTILADEMGLGKTVQTVIFVLSLYKEG 746

QY 319 -NLIPHVLAPLSTLNWREERATWAPQNMVYFCTAARAVIREHEEYVLSKQDKIKK 377
DB 747 HSKGFLVLSAPLSTIINWREEREMWAPDFVVTYTGDKESRSVIRENEF--SFEDNAIRS 804

QY 378 KKSQGISSESQKRIKFDVLTSTYEMINLDSAVLKPIKWECMIVDEGHRKKNKSKLFS 437
DB 805 GK--KVFRMKKEVQIKFHVLLTSYELITIDQALGSIENACLVDVDEAHLKKNQSKFFRV 862

QY 438 LTOYSSNHRILLTGTPPLQNNLDELFMHMFPLDAGKFGSLEEFQEBPKDINQSEQISRLHK 497
DB 497 LTOYSSNHRILLTGTPPLQNNLDELFMHMFPLDAGKFGSLEEFQEBPKDINQSEQISRLHK 497

```

```

DB 863 LNSYKIDYKLLLTGTPPLQNNLEELFHLNFLTPTPEFNNLEGLFLEBAFADISKEDQIKKLHD 922
QY 498 MLAPHLRRVKDKVMKMPKKELILRVDLSLQKEYYKAIFTRNYQVLTKKGA-QISL 556
DB 923 LLGPHMLRLKADVFKNPAKTELIVRVELSQMKYKFIILTRNFEALNSKGGNQVSL 982
QY 557 NNIMMELRKVCCHPYM--LEGVE-PVIHDANEAFKQLLESCGKLOLLDKMMVKLEKQGH 613
DB 983 LNIIMDLKCCNHPYLPFVAAVEAPLNGSYDSSLVKSGKMLLQKMLKLLRDEGHR 1042
QY 614 VLIYTOFOHMLDLLEDYCTHKWQYERIDGKVGGAERQIRIDRENAKSNKFCFLLSFA 673
DB 1043 VLIFSQMTKMLDLLEDYCTHKWQYERIDGKVGGAERQIRIDRENAKSNKFCFLLSFA 1102
QY 674 GGLGINLATADTVIYDSMDNPHADQAMARAHRLGQTNKVMYIRLINRGTIEERMMLT 733
DB 1103 GGLGINLATADTVIYDSMDNPHADQAMARAHRLGQTNKVMYIRLINRGTIEERMMLT 1162
QY 734 KKKVLEHLVVGK---LKTQINQBELDDIIRYSGKELFASE-----DD 774
DB 1163 KKKVLEHLVVGK---LKTQINQBELDDIIRYSGKELFASE-----DD 774
QY 775 EAGKSGK-----IHVDDAAIDKLDLDRDLVEAEVSDDEE 809
DB 1223 QSSKGNLAASAKKHGSGTTPPDNKNKVEDSVIHYDDAAISKLLDN-----QDATDTE 1277
QY 810 ---ENGFLKAFKANFEYIDENEAALAEQVAAESKSSAGNSDRASWEEILLKDKFELH 866
DB 1278 LONNVEYLSSFKVQYVREEDGVEEVEEIKQEE-----NVD-PDWKELRLHHYEQ 1331
QY 867 QAEELNALGKRKRKQL----VSIEEDDLAGLEVDSSDGSEY----EASTDGEAAG 918
DB 1332 QEDLARNLGKGRIRKQVYNDAEQDEB---WQDELSDNQSEYSIGSEDEDEFEERPE 1388
QY 919 GYQTCRRPYRK---GRNLEPTPLMEGEGRSFVLGPNQSORAIFVOTLWRYG----A 970
DB 1389 G-QSGRRSRLQKSDRDKPLP-PLLARVGNIEVLGPNARQKRAFLNAINMGWMPQDA 1446
QY 971 GNFMKKEFVPRLKOKTFEINEYGLFLKHAEEIDENSPTSDGVPKEGLRIEDVLVRI 1030
DB 1447 FNSHW--LYRDLRGKSEKEFRAYVSLFMRHLCEPGADGAETFADGVPREGLSRQHLVRI 1504
QY 1031 ALLIIVQKRVFVEDHPGKVPFPPSILRFRFGLRSGKIKWEEHDKIMIRAVLKHGGRWQ 1090
DB 1505 GWMSLVKRVKQVEFHVGKYSTPDLIPGEGPKSGEV----- 1542
QY 1091 AIVDKELGI-----QELICKELNPHISLSAEQAGLOGQSGSGSNPQAQTNQPGSVI 1146
DB 1543 -ISSDPNTFVPASPAPHLPLPGLPLD---KWEAQLGWYDE-----KDPGAQKPRP---- 1589
QY 1147 TGNNNASADGAQVNSMFYRDMQRLVFKVRKRVLLLEKAMNYEAYEYVGLGSSSIPT 1206
DB 1590 -----LEVALPALDRVESEDKH----- 1608
QY 1207 EPEAEPIADTVGVSFIEVDDEMGLPKTDPITSEIMGAADVNNOARVEIAHQNM 1266
DB 1609 ESPASKERARE-----ERPEETEKAPSPQLPREVL-----PEK 1644
QY 1267 CKLDENARESVOAVNN---QPPSTKYNESPRALKSINGINTILSITSDOSKSHEDDT 1323
DB 1645 EKILDKLSLHSHSGDSSELRPDTKAEKEPIETQNGD-----KBEDDEG 1692
QY 1324 KPDANNVWK---DFAETKPLRGVVDLNVV-EGEENIAEASGSV 1365
DB 1693 KKEDKKGKFKFMFIAD-----GGFTLHTLWQNEERAASVSGKI 1732

```

RESULT 4

Q81XZ5 PRELIMINARY; PRT; 1937 AA.

AC Q81XZ5; (TREMBLrel. 23, Created)

DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)

Db	1655	PMETEP--KGAADVKE-EKSAIDLTPIVVEDKEEKEEKEVML	1699
RESULT 5			
Q7WN3			
ID	Q7WN3	PRELIMINARY; PRT; 1893 AA.	
AC	Q7WN3;		
DT	01-JUN-2003 (TrEMBLrel. 24, Created)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Similar to chromodomain helicase DNA binding protein 4.		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;		
OC	Xenopodinae; Xenopus.		
OX	NCBI_TaxID=8355;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Embryo;		
RA	Klein S., Strausberg R.;		
RL	Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC046866; AAH46866.1; -		
DR	GO; GO:000785; C:chromatin; IEA.		
DR	GO; GO:0005634; C:nucleus; IEA.		
DR	GO; GO:0005524; F:ATP binding; IEA.		
DR	GO; GO:0003682; F:ATP dependent helicase activity; IEA.		
DR	GO; GO:0006333; P:chromatin binding; IEA.		
DR	GO; GO:0006333; P:chromatin assembly/disassembly; IEA.		
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.		
DR	InterPro; IPR000953; Chromo.		
DR	InterPro; IPR001410; DEAD.		
DR	InterPro; IPR002464; DEAH box.		
DR	InterPro; IPR001650; Helicase_C.		
DR	InterPro; IPR000330; SNF2_N.		
DR	InterPro; IPR001965; Znf_FHD.		
DR	Pfam; PF00271; helicase_C; 1.		
DR	Pfam; PF06628; PHD; 2.		
DR	Pfam; PF00176; SNF2_N; 1.		
DR	SMART; SM00298; CHROMO; 2.		
DR	SMART; SM00487; DEXDC; 1.		
DR	SMART; SM00490; HELIC; 1.		
DR	SMART; SM00249; PHD; 2.		
DR	PROSITE; PS00013; CHROMO_2; 2.		
DR	PROSITE; PS00690; DEAH ATP HELICASE; 1.		
DR	PROSITE; PS00016; ZF_PHD_2; 2.		
KW	Helicase.		
SQ	SEQUENCE. 1893 AA; 214687 MW; 5803A501C48BC520 CRC64;		
Query Match	28.9%; Score 2076; DB 13; Length 1893;		
Best Local Similarity	36.2%; Pred. No. 1.2e-107;		
Matches	526; Conservative 206; Mismatches 433; Indels 288; Gaps 47;		
QY	21	DSDDDDFVPKDRTTEQVEAIVRTAKENACQGESNLVSCNTCTVAFFIAKCLVPP 80	
Db	423	DLJDDAVGDPDES-----DHHMEFCRVCKDGGELLCCDVCPSYIHCLNP 470	
QY	81	KDASVENWRCEVCSP--LNEIDKILDCM-----RPTKSQSGSDAEPKPIF-----V 128	
Db	471	PEIPNGEWLCPRCTCPPLKGKIQLITWKVGQPPPTVSRPADSPDAAPLPLEGRPE 530	
QY	129	KOYLVMKGLSYLHCSWVPEKE-----FQAKYKSN-----HRLKTRV 165	
Db	531	REFFVKWAMSYLHCSWVTELQELHCVMFNRYQRKNDWDEPPAGDYGVDDEEKSRRK 590	
QY	166	NNFHROMESFNSEDFF--VAIRPEWTTVDRIIL-ACREEDGELEYLVKVELSYDECYWE 222	
Db	591	NKDPKYLE-----NEEKFYRIGKPEWMIHRLVNSVDKGYVHYLIKWRDLAYDOASWE 646	
QY	223	S-ESDLSFQNETQRF-----KDVNSRTRSKDVDHKK--NPRDFQ-----QFDHT 265	
Db	647	AEESDIQDIYKQGWNNHRLMCGDEGRPGKKVKVKURKLDPRDPTPAVDPTVKYDQ 706	
QY	266	PEFLK---GLLHPYQLEGLNLFRLFSWKQTHVILADEMGLGKTIQSIALLASLFEE--NL 320	
Db	707	FDYLDATGTLHPYQLEGLNLFRLFSWAQGTDTILADEMGLGKTVQTAFLVLSYKEGHSK 766	
QY	321	IPHLVIAPIASTLRNWEREFATWAPOMNVVMYFGTAQARAVIREHEFEYLSKQDKTKKKS 380	
Db	767	GFPLVSAPLSTIINWEREFEMWAPDMYVTVYVGDKDSRAVIRENEFESEGNAIRGKKAS 826	
QY	381	GQISSESKQKRIKFDVLLTSYEMINLDSAVLKPIKWECEMIVDEGHLKXKDSKLFSSITQ 440	
Db	827	RM----KKEASVKFHVLLTSYELITIDTAVLGSIDWACLVDDEAHLKXNQSFPFVLNG 882	
QY	441	YSSNRIILLTGTPIQNNLDLPMHFLDACKFGSLEFEQBFKDINOBEQISRLHMLA 500	
Db	883	YPLQHKLLLTGTPLQNNLEELFHLNFLTPTPERFNNLEGLFEEFADIAKEDQIKKLHMLG 942	
QY	501	PHLLRVKVDYMKDMPKKEILIRVDSLSSLOKEYKAIPTFNYQVL--TKGGAQISLNNI 559	
Db	943	PHMLRLKADVPKMPKSTELIVRVELSPMQKKYKFLITRNFETAINTRGNGNQVSLNV 1002	
QY	560	MMEFLKVCCHPYMLEGVBPVIHDANEAFKQ-----LLESCGKLQLLDKMMVKLRQ 610	
Db	1003	VMDLAKCCNHPVLF---PVA--AMEAPKMPNGMYDGSALIKGAGKLLFQKMLRKLDD 1056	
QY	611	GHRVLIYTFQFQHMDDLLEDYCTHKWKQYERIDGKVGGAERQIRDRFNAKSNKFCFLLS 670	
Db	1057	GHRVLIFSQMTKMLDLEDPMHEGKYKYEIDGGITGNMQEADIRFNAPGAQQFCFLLS 1116	
QY	671	TRAGLGIGNLATADTVIYDSMDNPHADLOAMARAHRLGOTNMKMIYRLINRGTTIERM 730	
Db	1117	TRAGLGIGNLATADTVIYDSMDNPHNDIQAFSAHRHIGQNRKVMYIRFVTRASVEERIT 1176	
QY	731	QUTKKKMLEHLVVGK---LKTQINIQBELDDIIRYGSKELFASDEDDAG--KSGK---- 781	
Db	1177	QVAKKKMMLTHLVVRPGLSGTSGMSKQELDDILLKFGTEELFKDTEGGENKEGEDISV 1236	
QY	782	IHYDDAAIDKLLDRDLVEAEVSDDEENGFLKAFKVANFEYIDENEAALAEARVAAE 841	
Db	1237	IHYDDKAIARLIDRNQDETEPEL--QLGMNYSLSFKVAQYVREENGDEEVVREIK 1295	
QY	842	SKSSAGNSDRASYWEELKDKFELHQAELNALGKRKSRKQL-----VSIEDDLA 897	
Db	1296	QEEVD-----PDYWEKLLRHYYEQQEDLARNLHGKRIKQVNVNDSQEDRD---WQD 1348	
QY	898	VSSDGESEYAESTDG--EAAQGVQGTGRRPYR---RKGRLNLEPTPLMEGERSFVLGF 953	
Db	1349	DQSDNQSDYSVASEGDEDFDERSEARRPNRKLNDKDXPLP--PLLARVGNIEVLGF 1407	
QY	954	NOSQRAIFVQTLIRYG-----AGNFDWKEFVPRLKQKTFEINEYGIILFKHIAEIDEN 1008	
Db	1408	NARQKAFALNIMRYGMPDQAFATQW--LVRLRGKSEKBFKAYVSLFMRHLCEPFGADG 1465	
QY	1009	SPTPSDGVPKGLRLIEDVLVRIALLILYQEKVKFVEDHPGKVPFVPSRILERPGLRSKI 1068	
Db	1466	AETFDAGVPRGLSRQHVLTRIGVMSLRKKVQEPF----- 1501	
QY	1069	WKEEHDKIMIRAVLKHGYSRQWAI VDDKELGTQELICKELNPPHLSLAAEQAGLQONG 1128	
Db	1502	-----HVNGRW-----SNP--ELAEAEENKASKID 1525	
QY	1129	SGGSNPGATQNPQNSVITGNNNSADGAQVNSMPYRDMORRLVEFKVRVLLLEKAWN 1188	
Db	1526	S--PSPKPTSTSGD--TQNTPTAPPASNEEDST-----EENNETTQPTPT----- 1555	
QY	1189	YEYAEYVYGLGSSSIPTPEEPEAEPKIADTVGVSFIEVDDEMLDGLPKTDPTITSEIMGA 1248	
Db	1556	-----KSEEVPKREETDNEVKPT-----EENNETTQPTPT-----OSP 1590	
QY	1249	AVDNNQARVEIAQHYNQMKLLDENARESQAQVYVNNQPPSTKVNESFRALKSINGNINTI 1308	
Db	1591	AAQTEKEKEEPVQ-----QERTSPPPPPPPPAPEPMDQ--KPEPEKVDV 1637	
QY	1309	LSITSDQSKSHEDDTKPDINNVM-----KDTABE--TKPLR-----GGV 1347	

```
Db 1638 ITL-----DDKKDDQEVTLONGETPKETAEBKTKKASAAAQRFMFNIADGGFT 1688
QY 1348 DINVV-EGENIA 1359
Db 1689 ELHSLWQNEERAA 1701

RESULT 6
QY410
AC QY410 PRELIMINARY; PRT; 2000 AA.
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Zinc-finger helicase.
GN HZFH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98351552; PubMed=9688266;
RA Aubry F., Mattei M.G., Galibert F.;
RT "Identification of a human 17p-located cDNA encoding a protein of the
RT Snf2-like helicase family.";
RL Eur. J. Biochem. 254:558-564(1998).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 "CHROMO" DOMAIN.
DR EMBL; U91543; AAC39923.1; -.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:008270; F:zinc ion binding; NAS.
DR GO; GO:0008355; P:regulation of transcription, DNA-dependent; NAS.
DR InterPro; IPR00953; Chromo.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002464; DEAD box.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR InterPro; IPR001965; Znf_PHD.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00385; chromo; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF0628; PHD; 2.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00298; CHROMO; 2.
DR SMART; SM00487; DEXDc; 1.
DR SMART; SM00490; HELICc; 1.
DR SMART; SM00249; PHD; 2.
DR SMART; SM00184; RING; 2.
DR PROSITE; PS00598; CHROMO_1; 1.
DR PROSITE; PS0013; CHROMO_2; 2.
DR PROSITE; PS00690; DEAD ATP HELICASE; 1.
DR PROSITE; PS0016; ZF PHD 2; 2.
KW ATP-binding; Helicase; Hydrolase; Nuclear protein.
SQ SEQUENCE 2000 AA; 226576 MW; B085A42A5AC7971F CRC64;

Query Match 28.7%; Score 2066; DB 4; Length 2000;
Best Local Similarity 33.5%; Pred. No. 4.6e-107;
Matches 520; Conservative 211; Mismatches 415; Indels 404; Gaps 41;

QY 52 COACGESTNLVSCNTCTAFHAKCLVPLKADASVENWRCPEVCVPL--NEIDKILDCMR 109
Db 459 CRVCKDGGELCCDAICISSYHCHLNPPLFDIPNGEWLCPRCTCPVLKGRVOKILHWRWG 518
QY 110 -----PTKSSQGGSSDA-EPKPI---FVKQYLVKWKGLSYLHCSWPVEKBFQAKSNH 159
Db 519 EPPVAVPAQADGPDVPPPPPLQGRSEREFVVKVGLSYWHCSWAKELQLEIHLVMY 578
QY 160 RLKTRVNNFRQWE-SFNNSDD-----FVAIRPEWTTVDRI-- 196
Db 579 RNYQRKNDMDPEPPLDYGSGDDGKDKRKVKVDPHYAEMEEKYRYFGIKPEWMTVRIIN 638
QY 197 ACREEDGELEYLVKVKELSYDCYW-ESESDISTFQEIQ---RFDKVNSTRRSKVDH 252
```

```
Db 639 HSDVKKGNHYLVKWRDLPYDOSTWEDEMINIPEYEEHKQSWRRELIMGEDPQPRKY 698
QY 253 KRNPRDFQ-----QFDHTPEFLK---GLLHPYQLEGFLNFRFSWSKQTHVI 295
Db 699 KKKKELQDGGPPSSPTNDPTVKYETQPRFITATGTLHMYQLEGFLNFRFSWAQCTDFI 758
QY 296 LADENGLCKTQSIALLASLPEENLI--PHLVIAPLSTLRNWEREFATWAPQVNVYFG 353
Db 759 LADENGLCKTIQTIIVFLSYLKEGHTKGFPLVSAFLSTIINWEREFQWMAPEYVYVYTG 818
QY 354 TAQARAVIREHEFYLS-----KDQKKIKKKSQISSSEKOKRIKFDVLLTSTYEMINLDSA 409
Db 819 DKDSRAIIRENEFSFEDNAIKGKKAFWK-----REAQVKFHLVLTSLTIDOA 870
QY 410 VLKPIKWCMIVDGHRLLKNKDSKLFSSLTQYSSNHRILLTGTPLQNNLDELFLMLHFLD 469
Db 871 ALGISRWACLVVDEAHLKNNQSKFFRVLYGKIDHKLLLTGTPLQNNLEELFHLNPLT 930
QY 470 AGKFGSLEEFQEEFKDINQEQIISLHKWLAPHLRLRVKKQVMDKMPKKEILRLVDLS 529
Db 931 PERFNNLEGFLEEFADISKEDQIKLHDLGLHMLRLKADVFKMPPAKTELIIVRVELSP 990
QY 530 LQKEYYKAIFTRNYOVLTKKGA-QISLNNIMMELRKVCCHPYMLEGVEPVITHDANEAPK 588
Db 991 MQKKYKYLITRNFALNSRGGGQVSLNIMMDLKCCNHPYLF---PVA--AMESPK 1044
QY 589 -----QLLESCGLQLLDKMMVKLEQGHRLVIYTFQHMLDLDLEDYCTHKWQYE 639
Db 1045 LPSGAYEGALIKSGKMLLQKMLKEQGHRLVFSQMTKMLDLEDYEGYKYE 1104
QY 640 RIDKVGGAERQIRIDRFNAKSNKPCFLSTLSTRAGGLGINLATATVYIYSDWNPHADL 699
Db 1105 RIDGITGALRQEAIDRFNAPCAQCFCLSTRAGGLGINLATATVYIFSDWNPHPNDI 1164
QY 700 QAMARAHGLGOTKVMYVRLNRTGIERRMQLTKKKVLHGLVVGK---LKTQINQEE 756
Db 1165 QAFSAHRIGQANKVMYIFRVTRASVEERITQVAKRKMLTHLVVRPGLSGKAGSWSKQE 1224
QY 757 LDDIIRYSGKELFASDDDEAGK---SGKIHYDDAAIDKLLDRDLVEABEVSVDDEENG 813
Db 1225 LDDILKFGTEELFKDNEGENEKEEDSSVHYDNEAIAELLDRNQDATEDTDV--QNMNEY 1282
QY 814 LKAFKVANFEYIDENEAALAEQARVAESKSAGNSDRASWHEELKKDFELHQAELNA 873
Db 1283 LSSFKAQYVVRREEDKIEIEREIIKQEE-----NVD-PDYWEKLLRHHEQQOEDLARN 1336
QY 874 LGKRRSRKQL--VSTEEDDLAGLDVSSDGESEYAEASTDGEAAGOGVQTRPYRRK 930
Db 1337 LGKGRVRVKQVYNDAQED-----QDNQSEYSGSEEDDFDRPGRKQSKQLRNE 1391
QY 931 GRDNLPTPLMEGEGRSFVLGFGNQSORAIFVQTLIRYQ-----AGNFDWKEFVPRLKQK 985
Db 1392 -KDXPLP-PLLARVGGNIEVLGFTNRQKAFNAVMRGMPQDAFTQW--LVRLDACK 1447
QY 986 TFEINEYGIILFKHIAEIDENSPTSDGVPKGLRIEDVLVRIALLILVOEKVKFVED 1045
Db 1448 TEKEFKAYVSLFMRHLCEPGADGETFADGVPRBGLSRQVLTIRIGVMSLVKKVQOEFEH 1507
QY 1046 HPG-----KPVFP----- 1053
Db 1508 INGRWSMPELMPDPSADSKRSRASSPTKTSPTTPEASATNSPCTSKPATPAPSKSGGI 1567
QY 1054 -----SRILRF----- 1060
Db 1568 RTPLEKEAEAEQEKPEKNSRIGEMKETEADAPSPASGLERLBPRTIPEDEVPGVGE 1627
QY 1061 ----PGLR----- 1064
Db 1628 MEPEPGYGRDKRSKATESTPCGERGEKPLDQGEHREPEGETGLGKREDVKGDRELAPG 1687
QY 1065 -----SGK---IWKEEHDKI 1076
```

```
Db 1688 PRDEPSRNGRREKTEKNTENPRFMFNADGGFTLHTLWQNEBERAAISSGKLNEIWHRRHDYV 1747
QY 1077 MTRAVLKHGVRQWQATVDDKELGQIBELICKELNFPHISLSAAEQAGLQONGSGGSPCA 1136
Db 1748 LIAGIVLHGVARQWDIQDAQFAI-----INEPF----- 1776
QY 1137 QTGNQFSGVITGNNASADGAQVNSMFYYRDMQRRVLFVFKRVLLLEKAMNYE---YAE 1193
Db 1777 KTEANKGNFLEMKV-----KFLARRFKLLEQALVIEQLRRA 1813
QY 1194 EYGLGSSSIPTEEPAPKPIADTVGVSFIEVDDMLDGLPKTDPTITSEEMGAVDNN 1253
Db 1814 AYLNLQEPAPFAMALHARFAEAECIAESHQHLSKESLAG-----NK 1855
QY 1254 QARVEIAQHYNQMKLLDENARE--SVQAVVNNQPP-STKNVESFRALKS 1300
Db 1856 PANAVLHKVNLQLELLSDMKADVTRLPATLSRIPPPIARLQMSERSILS 1905

RESULT 7
Q19815 PRELIMINARY; PRT; 1829 AA.
AC Q19815;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE F26F12.7 protein (LET-418).
GN F26F12.7 OR LET-418.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.:
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38 (1994).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Wilson R., Bentley D., Gattung S.;
RT "The sequence of C. elegans cosmid F26F12.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE FROM N.A.
RA von Zelewsky T., Palladino F., Brunschwig K., Hainal A., Mueller F.;
RT "The C. elegans Mi-2 chromatin-remodeling proteins function in vulval
RT cell fate determination.";
RL Development 0:0-0 (2000).
DR EMBL; U55373; AAC25894.1; -.
DR EMBL; AF308445; AAC29838.1; -.
QY 40 EAIVRTDAKEN--ACQAGESTNLVSNCTTVAFAKCLVPLPKDA--SVENWRCEC--V 94
Db 306 EYVKEAPKQNDFFCKICKETENLLCSCVCFHAYCIDPPLTEVPKEETWSCRCETV 365
QY 95 SPLNIDKILDCMR-----PTKSSEQSSD-----AEPKPIFKVLYKWKGLSYL 141
Db 366 KPEHKIEKILCWWRKEIPEYPLEAGKASSDAMLKPPRQWEPREFFVFWKYLSTW 425
QY 142 HCSWPEPEFQKAYKSNHRLKTRVNN-----FHRQMESFNNSDD-----FVAIR 186
Db 426 QCSWSEMLLEVFHFRMLILLYWRKNDSDAPPEFESVTSRHSNDNDPYKLRRFYQYGIK 485
QY 187 PEWTTVDRIACRE-EDGELEYLYVKYKELSYDECTWE--SESDISTFQNIQRP----- 237
Db 486 PEWQIHRIIINHQSAYAKSQDYLYKWKELSYDQATWERDDSDNIANYEEAIKYWQHRESK 545
QY 238 --KDVNSRTR-----SKDVHKNRPD-----FOQPDHTPEFLK---GLLHPY 276
Db 546 LNEIDIPKNVQKMIAXHREAKGLPPKEDEKKKKREKIDIRKKYEVQPDVTTETGGKLHPY 605
QY 277 QLEGLNLFKFSWKSQTHVILADEMGLGKTIQSIALLASLFEENLI--PHLVIAPLSTLRN 334
Db 606 QLEGLNLHCHWSNGTDAILADEMGLGKTVQSLTFYLSLMKGCHCKGPPFLIAAPLSTIIN 665
QY 335 WEREFPATWAPQNVNVYFGTAQARAVIRHEHYLSKQKKIKKKKSGQISSSKQKRIKF 394
Db 666 WEREABQWCPDFYVVTYVGLRDARVVLRREHFSFVEGAVRSGPKASKMKTTEN---MKF 721
QY 395 DVLTLSTYEMINLDSAVLKPDKWECHMIVDEGHRLLKKNKSKLFSLSLTQYSSNHRILLTGTP 454
Db 722 HVLLTSYETINMDKTLTSSIEWGLVVDHARLKNQSLFFKXNLNEYTHYRVLITGTPL 781
QY 455 QNNLDELFLMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHKMLAPHLRLRRYKDVMDK 514
Db 782 QNNLEELFHLNLFSLKERNQLEAPTAEFNEISKEDQIEKHLNLLGPHMLRLRLKADVLTG 841
QY 515 MPPKKELIURVLDLSSLQKQYKAI FTRNYQVLT--TKKGGAQISLNNIMBELRKCCHPYML 573
```

```
DR PIR; T34239; T34239.
DR WormPep; F26F12.7; CB17716.
DR GO; GO:000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0003682; F:chromatin binding; IEA.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR GO; GO:0006333; P:chromatin assembly/disassembly; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000953; Chromo.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002464; DEAH box.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR InterPro; IPR001965; Znf_PHD.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00385; Chromo; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00628; PHD; 2.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00298; CHROMO; 2.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR SMART; SM00249; PHD; 2.
DR SMART; SM00184; RING; 2.
DR PROSITE; PS00013; CHROMO 2; 2.
DR PROSITE; PS00690; DEAH ATP HELICASE; 1.
DR PROSITE; PS50016; ZF_PHD 2; 2.
KW ATP-binding; Helicase; Hydrolase.
SQ SEQUENCE 1829 AA; 209150 MW; 1A887E990C63B661 CRC64;

Query Match 28.5%; Score 2048.5; DB 5; Length 1829;
Best Local Similarity 32.9%; Pred. No. 3.9e-106;
Matches 518; Conservative 253; Mismatches 474; Indels 331; Gaps 44;

QY 40 EAIVRTDAKEN--ACQAGESTNLVSNCTTVAFAKCLVPLPKDA--SVENWRCEC--V 94
Db 306 EYVKEAPKQNDFFCKICKETENLLCSCVCFHAYCIDPPLTEVPKEETWSCRCETV 365
QY 95 SPLNIDKILDCMR-----PTKSSEQSSD-----AEPKPIFKVLYKWKGLSYL 141
Db 366 KPEHKIEKILCWWRKEIPEYPLEAGKASSDAMLKPPRQWEPREFFVFWKYLSTW 425
QY 142 HCSWPEPEFQKAYKSNHRLKTRVNN-----FHRQMESFNNSDD-----FVAIR 186
Db 426 QCSWSEMLLEVFHFRMLILLYWRKNDSDAPPEFESVTSRHSNDNDPYKLRRFYQYGIK 485
QY 187 PEWTTVDRIACRE-EDGELEYLYVKYKELSYDECTWE--SESDISTFQNIQRP----- 237
Db 486 PEWQIHRIIINHQSAYAKSQDYLYKWKELSYDQATWERDDSDNIANYEEAIKYWQHRESK 545
QY 238 --KDVNSRTR-----SKDVHKNRPD-----FOQPDHTPEFLK---GLLHPY 276
Db 546 LNEIDIPKNVQKMIAXHREAKGLPPKEDEKKKKREKIDIRKKYEVQPDVTTETGGKLHPY 605
QY 277 QLEGLNLFKFSWKSQTHVILADEMGLGKTIQSIALLASLFEENLI--PHLVIAPLSTLRN 334
Db 606 QLEGLNLHCHWSNGTDAILADEMGLGKTVQSLTFYLSLMKGCHCKGPPFLIAAPLSTIIN 665
QY 335 WEREFPATWAPQNVNVYFGTAQARAVIRHEHYLSKQKKIKKKKSGQISSSKQKRIKF 394
Db 666 WEREABQWCPDFYVVTYVGLRDARVVLRREHFSFVEGAVRSGPKASKMKTTEN---MKF 721
QY 395 DVLTLSTYEMINLDSAVLKPDKWECHMIVDEGHRLLKKNKSKLFSLSLTQYSSNHRILLTGTP 454
Db 722 HVLLTSYETINMDKTLTSSIEWGLVVDHARLKNQSLFFKXNLNEYTHYRVLITGTPL 781
QY 455 QNNLDELFLMHFLDAGKFGSLEEFQEEFKDINQEEQISRLHKMLAPHLRLRRYKDVMDK 514
Db 782 QNNLEELFHLNLFSLKERNQLEAPTAEFNEISKEDQIEKHLNLLGPHMLRLRLKADVLTG 841
QY 515 MPPKKELIURVLDLSSLQKQYKAI FTRNYQVLT--TKKGGAQISLNNIMBELRKCCHPYML 573
```



```
QY 225 -SDISTFQNEIQRFKDV-----NSRTRRSKQVDHKNR-----DFQQPDHTPEFLK 270
DB 406 IQDYDLFKQSYWNHRELMEGRPGKKLKKYKRLKLERPPPTPTVPTVKYERQPEYLD 465
QY 271 ---GLLHPYQLEGLNFRSWSKQTHVILADEMGLGKTIQSTALLASLPEE--NLIPHVL 325
DB 466 ATCGTLHPYQMEGLNWLRSWAQGTDTILADEMGLGKTVQTAFLVSLYKEGSHGKPFVL 525
QY 326 IAPLSTLRNWEREFATWAPQNMVVMYFGTAQARAVIREHEFYLSKDOKKIKKKKGQISSL 385
DB 526 SAPLSTIINWEREFENWAPDMVVTYVGDKDSRAIIRENEFPEDNAIRGGKASRM---582
QY 386 ESKQRIKFDVLLTSYEMINLDSAVLKPIKWCMIIVDEGHRLLKNKDSKLFSSITQYSSNH 445
DB 583 -KKEASVKFHLVTSYELITIDMAILGSDWACLIVDEAHRLLKNQSKFFRVLNGYSLOH 641
QY 446 RILLTGTPLQNNLDELFLMLHFLDAGKFGSLBEFOEPKIDINOEQISRLHKLMLAPHLRL 505
DB 642 KULLTGTPLQNNLEELPHLNLFLTPERFNLGFLFEFADIAXEDQIKKLHMLGPHMLR 701
QY 506 RVKQVMDKMPKPKELILRVDSLSSLOKYYKAITFRNYQVLTKGGA-QISLNNIMMELR 564
DB 702 RLKADVKMPSKTELIVRVLSLSPMKQKYKYLITENPEALNARGGQVSLVNVMDLK 761
QY 565 KVCHPMLGVEPVTIHDANEAFKQ-----LLESCGKLQLLDKMMVKLKEGHRVL 615
DB 762 KCCNHPFLF---PVA--AMEAPKPMNGMYDGSALIRASGKLLLKQMLKNIKEGHRVL 815
QY 616 IYTOFQMLDLLEDYCTHKKWOYERIDGKVGGAERQIRIDRENKSNKFCFLSTRAGG 675
DB 816 IFSQMTKMLDLLEDFLUEHGKYKFERIDGGITGNMRQEAIDRENAPGAQFCFLSTRAGG 875
QY 676 LGINLATADTVIYDSMDNPHADLQAMARAHRLGQTNKVMYIRLINRGTIERMMQLTKK 735
DB 876 LGINLATADTVIYDSMDNPHNDIQAFSRAHRIGQNKVMYIRFVTRASVEERITQVAKK 935
QY 736 KMWLEHLVVGK---LKTQNIQEBLDDIIRYSGKELFASDEBAGSKG-----IHYD 785
DB 936 KMWLTHLVVRPGLSKTSMSKQELDDILKFGTEELFKDEATDGGDNKEGSDSSVIHYD 995
QY 786 DAAIDKLDRDLVEAREVSDDEENGFLKAFKVANFEYIDENEAAALEAQR 837
DB 996 DRAIEKLLDRNQDETDTEL--CGMNEYLSFPKVAQIVVREEMGEEMEEVEVER 1045
```

RESULT 10

```
O42142
ID O42142 PRELIMINARY; PRT; 1808 AA.
AC O42142;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Chromo-helicase-DNA-binding on the Z chromosome protein.
GN CHD-Z.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97473516; PubMed=9332370;
RA Griffiths R., Korn R.M.;
RT "A CHD1 gene is Z chromosome linked in the chicken Gallus
RT domesticus";
RL Gene 197:225-229(1997).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 2 'CHROMO' DOMAINS.
DR EMBL; AF004397; AAC60282.1; -.
DR HSSP; P23197; 1A0.
DR GO; GO:0000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
```

```
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
DR GO; GO:0003682; F:chromatin binding; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006333; F:chromatin assembly/disassembly; IEA.
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000953; Chromo.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase C.
DR InterPro; IPR001005; Myb DNA Binding.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00385; chromo; 2.
DR Pfam; PF00271; helicase C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00298; CHROMO; 2.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR PROSITE; PS00598; CHROMO_1; 2.
DR PROSITE; PS00013; CHROMO_2; 2.
DR PROSITE; PS00334; MYB_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW ATP-binding; DNA-binding; Helicase; Hydrolase; Nuclear protein.
SQ SEQUENCE 1808 AA; 208399 MW; 93A07FF7B401B5D8 CRC64;

Query Match 22.0%; Score 1578; DB 13; Length 1808;
Best Local Similarity 29.7%; Pred. No. 1.1e-79;
Matches 444; Conservative 255; Mismatches 457; Indels 340; Gaps 50;

QY 7 RLRTSRDRKPV-----YNLDDS-----DDDFVPVKDR-----TFPQVEALVRTDAKEN 50
DB 194 RIKPSGKKGKQKQKRLQDSEEDDDDEDYDKGRSRRQATVNVYSKAAEE--TKTDS--DD 251
QY 51 ACQACGESTNLVSCNTCTYAHAKCLVPPLKXASVENWRCPECVSPLNEIDLCEMRP 110
DB 252 LLEVCGED-----VPQTEDEFET-----IEKFM--SRI 279
QY 111 TKSSEQSGSS-----DAEPKPIPVK-----QYLVKWKGLSYLHCSWVPEKEFQKAY 155
DB 280 GRKGATGASTIYAVEADGDPNAGFEKSKELGEIQYLIKWKGWSHIHNWETEETLK---336
QY 156 KSNHRLKTRVNNFHRQ-----MESFNSEDDFVAIRPEWTVVDRILACRE 200
DB 337 QQNVKGMNKLNDYKKKQDETFRWLKNASPEDVEYYNCOQELTDLLHKQYQIVERIASHN 396
QY 201 ED---GELEYLVKVELSYDECYMESESDIS--TFQNEIQRFKQVNV--SRTRSKQVD--HCR 254
DB 397 QKSAAGYPDYCKWQGLPYSECSWEDGALIAKKQFQARIDEFYFRNQSKITTPFKCKVLKQ 456
QY 255 NPROFPQQPDHTPEFLKG---LLHPYQLEGLNFRFSWSKQTHVILADEMGLGKTIQSTIA 310
DB 457 RPR-FVALKKQPSYIGGHESLELRDYQLNGLNLWLAHSAWCKNSCILADEMGLGKTIQSTIS 515
QY 311 LLASLFEENLI--PHLVIAPISTLRNWEREFATWAPQNMVVMYFGTAQARAVIREHEFVL 368
DB 516 FLNVLFEHQYLGPPFLRVPLSTLTSMQREIQTWAPQNMVAVYLGDTISRNMRIRTHEW--573
QY 369 SKDQKKIKKKSGQISSSESKOKRIKFDVLLTSYEMINLDSAVLKPIKWCMIIVDEGHRILK 428
DB 574 -----MHPQTRKLFKNILLTYETILLKDKGFLGGLNNAFVGVDSAHRLK 617
QY 429 NKDSKLFSSLTQYSSNHRILLTGTPLQNNLDELFLMLHFLDAGKFGSLBEFOEPKIDINO 488
DB 618 NDDSLLYRTLLDFKSNHRLILLTGTPLQNSLAKSLHLHFIMPEKFPSSWEDFEHEHGK--GR 676
QY 489 EEQISRLHKMLAPHLRLRRVKDKVMKMPKPKELILRVDSLSSLOKYYKAITFRNYQVLT 548
DB 677 EYGVASLHKELEPFLRRVKDKVEKSLPAKVEQILRMEMSAIQKYKWLITRNYKALSK 736
QY 549 -KGGAQISLNNIMMELRVCCCHPYMLEGVPEP-VIHDANEAFKQLLESCKQLQLDKMMVK 606
```


Db 1338 SDEDEPK-REYTAAGLALKEKYEKLARQKERI 1370

RESULT 12

Q9S141 ID Q9S141 PRELIMINARY; PRT; 1738 AA.

AC Q9S141

DT 01-MAY-2000 (TRENBLrel. 13, Created)

DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)

DE SNF2 family (CHD1 subfamily) chromodomain protein.

GN AT2G13370.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=20083487; PubMed=10617197;

RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,

RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,

RA Buell C.R., Ketchum K.A., Lee J.J., Renning C.M., Koo H., Moffat K.S.,

RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,

RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,

RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,

RA Salzberg S.L., Fraser C.M., Venter J.C.;

RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis

RT thaliana."

RL Nature 402:761-768 (1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RA Lin X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC007209; RAD28668.1; --

DR PIR; C84507; C84507.

DR GO; GO:0000785; C:chromatin; IEA.

DR GO; GO:0005634; C:nucleus; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.

DR GO; GO:0003682; F:chromatin binding; IEA.

DR GO; GO:0016787; F:hydrolase activity; IEA.

DR GO; GO:0006333; P:chromatin assembly/disassembly; IEA.

DR InterPro; IPR000953; Chromo.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR001650; Helicase_C.

DR InterPro; IPR000330; SNF2_N.

DR Pfam; PF00385; chromo; 1.

DR Pfam; PF00271; helicase_C; 1.

DR Pfam; PF00176; SNF2_N; 1.

DR SMART; SM00298; CHROMO; 1.

DR SMART; SM00487; DEXDC; 1.

DR SMART; SM00490; HELICC; 1.

DR PROSITE; PS00013; CHROMO 2; 2.

KW ATP-binding; Helicase; Hydrolase.

SQ SEQUENCE 1738 AA; 198696 MW; CE50069DD8C3B6CD CRC64;

Query Match 21.8%; Score 1565.5; DB 10; Length 1738;

Best Local Similarity 30.7%; Pred. No. 5.1e-79;

Matches 462; Conservative 237; Mismatches 485; Indels 321; Gaps 50;

QY 11 RSDRPVNLDDSDDDFPVKDRTPEQVEAVR--TDAKENACQACGESTNLV---SC 64

DB 329 RQKRKTSYQDDSEED-----SENDNDEGRSLARCTTLRQNN-----GRSTNTIQSSEV 380

QY 65 NTCTYAFHAKCLVPPLKLDASVENRCECVSPINE-----IDKILDCMRPTKS 113

DB 381 RSSTRSVRKVSVESESDSDIDGK-----NRKNQKDDIEEDADVIEKVLWHQLKGME 435

QY 114 SQGGSSDAPKPIFVKQ-----YLVKWKGLSYLHCSWVPEKEFK----- 153

Db 436 DVQ-TNNKSTVPVLVSQLFDTPEPDWNEFLIKWKQSHLHCQWKTLSDLQNLSGFKKVL 494

154 --AYKSNHRLKTRVNNFHRQMESNNSEDDFVAIRPEWTTVDRIILACR-EDG-----BLE 206

495 NYTKKVTETIRYRTALSRREEIVNDVSKEMOLDIKQNSQVERIADIRISKDGLGVVPE 554

207 YLVKYSYDECYWESSEDISTFQNEIQREKDVN-SRTRSKVDVHKR----- 254

555 YLVKQGLSYAEATQEKDQVDFAFQVADFYKAREVSIAGQKMWQOQTGKGNSFSN 614

255 -----NPRDFQFDHTPEFL-KGLLHPYQLEGLNFLRFSWQTHVILADEMGLGKTI 306

615 AELWLLFSVASLRKLDEQPEWLGTLRDYQLEGLNPLVNSWLNDDTNVILADEMGLGKTV 674

307 QSIALLASLPENLI--PHLVIAPLSTLRNWEREPATWAPQWVVMYFTGAQARAVIREH 364

675 QSVNMLGFLQNTQITPGPFLVVPVPLSTLANWAKEFRKWLPGMNIIVYVGTASREV---- 730

365 EFYLSKQKTKKKKSGQISSESQKRIKFDVLLTSYEMINLDSAVLKPIKWECMIVDEG 424

731 -----RNKTNVDVHKGRP--IKFNALLTYEVLVKDKXAVLSKIKWILYLVNDEA 776

425 HRLNKDSKLPSSLTQXSNHRILLTGTPLQNNLDELFLMHFLDAGKFGSLSEQEFBK 484

777 HRLKNSERQALYTALLEFSTKKNLLITGTPQNSVEELWALLHFLDPGFKNKDEFEVNYK 836

485 DIN--QBEQIRLHKMLAPHLLRRVKDKVMKMPKELIIRVLSLSLQKEYKAIPTRN 542

837 NLSSFNESELANHLELRPHILRRVIKVEKSLPPKIERILRVBMSPLQKQYKWIILRN 896

543 YQVLTG-KGGAQISINIMMELRKVCCHPYMLE---GVEPVIHDANEAFQQLSECKGL 597

897 FHDLNKGVGRGNQVSLNIVVELKKCCNHPFLFESADHGYGGDIND-NSKLKILIUSSKL 955

598 QLLDKMVMKKEQGRVLIYTFQFQHMLDLLEDYCTHKKWQYERIDGKVGGAERQIRIDRF 657

956 VILDKLLVLRRETKRVLIYFSQVMRLDILAEYLSLRCFGFQRLDGGSTKASLRQAMDF 1015

658 NAKSNKFCPLLSTRAAGLGINLATADTVIYDSWNPADLQAMARHRLGQTNKMIY 717

1016 NAPASDDFCFLSTRAAGLGINLATADTVIYFDSWNPQNDLQAMSAHRIQGEQVWNIY 1075

718 RLNRGTTEERMMLTKKMWLEHLVCKL-----KTONIQEELDDIIRVGS 765

1076 RFVTSKSVESIEILERRAKRMVLDHLVIOKLNAEGLREKRETKKGSFNDKNELSAILRGA 1135

766 KELFASE--DDEAGSGKTHYDADAIDKLLDRDLVEAEVSVDDDEENGFLKAFKVANFEY 824

1136 EELPKEDKNDESKKLLSMD---IDELER--AEQVEEKHTDETEHELLGAFKVANF-- 1188

825 IDENEAALAEQVAEAKSSAGNSDRASVYWEELLKKFELHQAFELNALGKR-KRSRKQ 883

1189 -----CNAEDDGSFWSRWIKPD-SVVTAEE--ALAPRAARNTKS 1224

884 LVSIEEDDLAGLEDVSSDGESEYAESTDGEAAGGVOTGRPPYRKRGRDNLPTPLMEG 943

1225 YVDPSPHDPRTSKR--KKKGSEPP-----HTERSQKRKRKTEYFVPSTPLLEG 1269

944 -----EGRSFRVLGNOSORAFVQTLRMRYGAGNPDWKEFVPRLLKQKTFEINEYGIPL 998

1270 TSAQVRGWSYGNLPKRDAQR--FYRTVNMKFGNN-----Qm 1303

999 KHIAEEDENSPTFSDGVPKEGLRIEDVLVRIALLILVOEKVKF-----VEDHPGKP 1050

1304 ACIAEEV-----GGVVEAAPAAQVELFDALIDGCKESVETGNFEPKPGVLDFGVP 1355

1051 VPPRILRRPFLR-----SGKIWEEDHKIMIRAVL 1082

1356 VKANELLRVQGLQLLSKRISRYNDPISQFRVLSYLKPSNWSKCGMNIIDARLLGL 1415

1083 KHGVRQWAIIVDDKELGIGQELICK-ELNFPHISLSAAEQAGLQGGGSGGSPGQATNQN 1141

1416 YHGFQNWBEKIRLDSLSGLTKKIAPVELQHHETFL-----PRAPNLKE 1457

GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000953; DEAD.
 DR InterPro; IPR001410; IPR001410; DEAD.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR001005; Myb DNA Binding.
 DR InterPro; IPR006025; Pept_M_Zn_BS.
 DR InterPro; IPR000330; SNF2_N.
 DR Pfam; PF00385; chromo; 2.
 DR Pfam; PF00271; helicase C; 1.
 DR Pfam; PF00176; SNF2_N; 1.
 DR SMART; SM00298; CHROMO; 2.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELIC; 1.
 DR PROSITE; PS00598; CHROMO_1; 2.
 DR PROSITE; PS00113; CHROMO_2; 2.
 DR PROSITE; PS00334; MYB_2; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW DNA-binding; Helicase.
 SQ SSQUENCE 1786 AA; 205974 MW; 23361A2C1A9FDF4A CRC64;
 Query Match 21.6%; Score 1553.5; DB 13; Length 1786;
 Best Local Similarity 29.4%; Pred. No. 2.5e-78;
 Matches 442; Conservative 247; Mismatches 427; Indels 385; Gaps 53;
 QY 20 LDSDDDDDFVFKD-----TFQVQVAVRTDAKENACQAGESTNLVSCNCTYAFH 72
 DB 201 LDSSDDDDYKGRSQRQATVNSYKEAE--TKTDS--DDLLEVCGED----- 244
 QY 73 AKCLVPLKDAVENWRCPECVSPLEIDKILDCENRPTKSSBOGS-----SDAEPK 124
 DB 245 ----VPQDEDEFET-----IEKFM--SRIGRKGATGAATTIYAVEADGDPN 286
 QY 125 PIVK-----OVLVKMGLSYLHCSWVPEKFOKAYKSNHRLKTRVNNFHRO----- 171
 DB 287 TGEKSEKPEGSIQYLKWKGSWHTWETEELK---QONVKMKLNNYKKQDQETR 343
 QY 172 -----MESFNSEDFFVAIRPEWTTVDRIACREED---GELEYLVKYKELSYDEC 219
 DB 344 WLKNSPEDVEYVNCQELTDDLHKQVQIVERIIAHSNQKSAAGYDYYCKWQGLPYSEC 403
 QY 220 WYSESDIS--TFQNETQRFQDVN--STRRSKQVD--HKNRPDPOQDHTPEFLKG---L 272
 DB 404 SWEDGALIAKKFQTCIDEYFSRNSQKTPFKCKILKQRP--FVALKQPSYTGHEGLE 462
 QY 273 LHYPQLEGFLNFRFSKQTHVLADBMGLGKTIQSIALLASLFEENLI--PHLVAPLS 330
 DB 463 LRDIYQNLNWLHSHCKGNSCILADBMGLGKTIQISFLNHLFHEHQIYGFLLVPLS 522
 QY 331 TLNRWEREFATWAPQNVVMYFCTAQARAVIREHEFYLSKDQKKIKKKXSGQISSSKQK 390
 DB 523 TLTWQREIQTWAPQNVAVVYLGDTISRNIRTHEW-----MHPQTK 564
 QY 391 RIKFVLLTSYEMINLDSAVLKPIKMECIVDGHRLKNKDKLFSLLTQYSSNHRILLT 450
 DB 565 RLKFNILLTTEILLKDKSPGLGNNVFTGVDEAHLKNDSDLLYKTLIDFKSNHRLIT 624
 QY 451 GTPQNQLDELFWLHFDAGKGSLEEFQEFKINDQEEQISLHKMLAPHLLRRVKKD 510
 DB 625 GTPQLNSKLKLSLLHFIIMPEKSSWEDFEEHGK--GREFGYASLHKELEPFLLRVKKD 583
 QY 511 VMKDMPKPELILRLVDLSLSQKEYKAIPTFRNYQVLTG--KGAQOISLNNIMBELRVCC 569
 DB 684 VEKSLPAKVQBIURMMSALQKYKWLITRNYKALSKGSGTSGTFLNIMBELKCCNH 743
 QY 570 PYMLEGVEP--VIHDANEAFQLESCKGLQLLDKMMVKLEQGHVLIYTFQFHMLDILE 628
 DB 744 CYLIKPPDDNEFYNQKEALQHLIRSSGKLLDKLIRLRECNRVLIIFSQVMVMDILA 803
 QY 629 DYCTHKWQYERIDGKVGGAERQIRIDRFNAKSNKFCFLLSFRAGLGGINLATADTVII 688
 DB 804 EYLKYQFPQRLDGSIKGLRQALDHFNAESGDFCFLLSFRAGLGGINLASATTVI 863
 QY 689 YDSDNPHADLQAMARHLGQTNKVMYRLINRGTIERMMQITKKKMWLEHLVVGKUK 748

DB 864 FDSWNPNQDLOAQAHRIGQKQVNTYRLVTKASVEEDILERAKKXKVLHVLQIRMD 923
 QY 749 TQN-----INQELDDIIRYSGSKELFASDEDDAGKSGKIHYDDAIDKLL 793
 DB 924 TTGKTVLHTGSTPSSSTPFNKEELSAILKFGAEELFKPEGEEREP-----QEMDIDEL 978
 QY 794 DRDLVEAEE---VSVDDEENGFLKAFKANFEYIDENEAALAEQRAVAESKSSAGNSD 850
 DB 979 KRAETRENEPGLTVGDE---LLSQFKVANFSNMDDED-----IELEPEQNLRN-- 1024
 QY 851 RASYWEEL-----LKDQFELHQAELNALGKRKRKRQKLVLSIEDDLAGLEVSSQG 902
 DB 1025 ----WEEIPEVQRRRIEERERQKELEIYMLPRNRCAKQ-----ISFNG 1066
 QY 903 DES-----YEAESTDGEAAGQVGTGRPRVRRKGRONLEPTPLMEGSGRFRVLGQ 955
 DB 1067 SEGKHSRRRRYSGSDS-----SVSEKRP--KKGRRPTIPRE-----NKGFS 1110
 QY 956 SQRAIFVOTLMRYGAGNDFWKEFVPR---LKQKTFEEINEYIGILFKHIAEIDENS--- 1009
 DB 1111 ASIRFIKSYKFG--GPLERLDAIARDAELVDKSIDLRLRGLGELVHNGCIKALKDNSFQ 1169
 QY 1010 -----PTFSDGVKPEGLRIEDVLRIALLILVQEKVKFVEDHPGKVPSPRIL 1057
 DB 1170 ERAGGRFGKVGKPTF-----RISGVQVNAKLVISSHEELAPL--HKSIAPDPEE-- 1216
 QY 1058 ERFPGLRSGKI-----NKEEHDKIMIRAVLKHGGRWQAVDDKELGIGQELI 1104
 DB 1217 -----RKRYVTPCHTKAAHPDIDMGKEDDSNLLIGIYEGYGSWEMIKWDPLSLTKI 1270
 QY 1105 CKELNFPHISLSAAEQAGLQGGQSGGNPGAQTNQFNSVITGNNNASADGAQVNSMPY 1164
 DB 1271 LPD-----DPKPKQA----- 1281
 QY 1165 YDMQRLVFEVKRVLLEKAMNVEYAEYGLGSSSIPTEPEAEPKIADTVGVSPFI 1224
 DB 1282 -KQLOTADYDLIK---LLNKDLARKEVQRLTGAGNS---KRKAKRSK- 1321
 QY 1225 EVDDEMLDGLPTDPTITSEEINGAAVDNNOARVEIAQHYNOMKLLDENARESVOQVNN 1284
 DB 1322 -----KNRA-----TKAAKKEEIKSDS----- 1339
 QY 1285 QPSTKVNESFRALKSINGNINTILSITSDQS--KSHEDDTKPDLLN-----VEMKDTA 1336
 DB 1340 SFPSEKSDDD-----GYNEDEIVSVKCPHKKIKAEENEKEPEPNTGIKMEAVEKRETR 1392
 QY 1337 E-----ETKPLRGVVDLNVVEGENIAEAS--GSVDYKMBEA---KEEEKPKQNV 1382
 DB 1393 EKENRHLKREKKEKEDKKEFNKIKERESKLKSTQKEKEKKEKANKELSENKENKK 1452
 QY 1383 V 1383
 DB 1453 V 1453
 RESULT 15
 Q800S6 PRELIMINARY; PRT; 1806 AA.
 ID Q800S6
 AC Q800S6;
 DT 01-JUN-2003 (Tremblrel. 24, Created)
 DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Chromo-helicase DNA-binding protein.
 GN CHDZ.
 OS Poephila guttata (Zebra finch) (Taeniopygia guttata).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Passeriformes; Estrildidae;
 OC Estrildinae; Taeniopygia.
 OX NCBI_TaxID=59729;
 RN [1]
 RA Agate R.J., Arnold A.P.;

RT "Sex Differences in Structure and Expression of the Sex Chromosome
 RL Genes CHD2 and CHD9 in Zebra Finches."; to the EMBL/GenBank/DBJ databases.
 DR Submitter (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY217131; AA061783.1; -
 DR GO; GO:0000785; C:chromatin; IEA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0003682; F:chromatin binding; IEA.
 DR GO; GO:0008237; F:metalloproteinase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0006333; P:chromatin assembly/disassembly; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000953; Chromo.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR001650; Helicase C.
 DR InterPro; IPR001005; Myb DNA binding.
 DR InterPro; IPR006025; Pept M_Zn_BS.
 DR InterPro; IPR000330; SNF2_N.
 DR Pfam; PF00385; Chromo; 2.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00176; SNF2_N; 1.
 DR SMART; SM00298; CHROMO; 2.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELIC; 1.
 DR PROSITE; PS00598; CHROMO_1; 2.
 DR PROSITE; PS0013; CHROMO_2; 2.
 DR PROSITE; PS00334; MYB_2; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW DNA-binding; Helicase.
 SQ SEQUENCE 1806 AA; 208045 MW; 6BC935637185FACO CRC64;
 Query Match 21.6%; Score 1551; DB 13; Length 1806;
 Best Local Similarity 29.5%; Pred. No. 3.5e-78;
 Matches 439; Conservative 247; Mismatches 475; Indels 328; Gaps 45;
 QY 7 RLRIISDRKEV---YNLDSDDDFVKK-----DRTFQVEAIVRTDAKE----- 49
 DB 193 RIKPKSGKSAGPKKQLDSEDDDDDEEDDDYDKGRSRQATVNISYKEABETKD 252
 QY 50 --NACQAGESTNLVSCNTCTAFHAKCLVPLKDAENVRCPECVSLNIDKILDC 107
 DB 253 SDDLLEVCED-----VPQDEDEFET-----IEKPM-- 280
 QY 108 MRPTKSEOGS-----SDAEPKPIFK-----QYLVKWKGLSYLHCSWVPEKEFQ 152
 DB 281 SRVGRKGATGAATTIVAVEADGDPNAGFEKSEPAEVQYLIKWKGWSHIHTWETEELK 340
 QY 153 KAYKSHRLKTRVNNFRQ-----MESFNSEDDFVAIRPEWTTVDRIILA 197
 DB 341 ---QQNVKGMKLDNYKKDKQETKRWLNKASPEDVEYNNCQBELTDDLHKQYQIVERITA 397
 QY 198 CREED--GELEVLVKYKELSDYECYWESESDIS--TFONEIQRFKDVN--SRTRRSKQVD- 251
 DB 398 HSNQKSAAGVPDYCKWQGLPYSECWEGDGLIAKFKQARIIDYFERNQSKTTPFKDCKV 457
 QY 252 HKRNPDRFOQDHTPEFLKGL-----LHPYQLEGFLNFRFSWQKTHVLADENGLGKTIQ 307
 DB 458 LKQRP-R-FVALKKQPSYIGGHESLGRDYQLNGLNLAHWSCKNGSCILADENGLGKTIQ 516
 QY 308 STALLASLFEENLI--PHLVIAPLSTLRNWEREFATQPMQNMVMTYFGTAQARAVIREHE 365
 DB 517 TISFLNLFHEHQLYGFLLVPLVPLSTLSQREIQTRAPQMNNAVYLGDITSRNMIRTHE 576
 QY 366 FYLSKDQKIKKKKSQISSESKQRIKFDVLLTSYEMINLDSAVLKPKWCMIWDEGH 425
 DB 577 W-----MHPQKRLKFNILLTYTTEILLKOKSFLGGLNWFVIGVDEAH 618
 QY 426 RLKNDKSKLFSLSITQSSNRIILLTGTPLQNNLDELFLMLHFLDAGKFGSLBEFQEPK 485
 DB 619 RLKNDLSLLYKTLIDFKSNHRLITGTPLQNSLKLWLSLLHFMPEKFSWEDFEHCK 678
 QY 486 INOEQIISLHKMLAPHLRLRVKVDVNMKMPKKEILILRVLSSLOQKEYKAIFTRNQV 545

Search completed: September 14, 2004, 02:07:47
 Job time : 177 secs

DB 679 -GREYGYASLHKELEPFLLRRVKDDVEKSLPAKVEQILRMESALQKQYKWLITGNKYA 737
 QY 546 LTK-KGGAQISLNNIMMELRKVCCHPYMLEGVEP-VIHDANEAFKOLLESCKQLLDLDM 603
 DB 738 LSKSGKSGTSGFLNIMMELKCCNHCYLIKPPDDNEFYNKQELQHLIRSSCKLILLDKL 797
 QY 604 MYLKEQGHVLIYTOFOHMLDLLEDYCTHKKWQYERIDGKVGGAERQIRIDRFNAKSN 663
 DB 798 LIRLRERGNRVLIIFSQVMRLDILAEYLYKYRQFPFQRLDGSIKGLRQALDHFNAEGSE 857
 QY 664 KFCPLSTRAGGLINLATADTVIYSDMNPADLQAMARAHRLGQTNKVMYVILNNG 723
 DB 858 DFCPLSTRAGGLINLASADTVIFSDMNPONDLOAQAHRIRGQKQVNIYILVYKG 917
 QY 724 TIEERMQLTKKMWLEHLVVGKLTQN-----INOEBLDDIIRYGSKE 768
 DB 918 SVEEDILERAQKQWLDHLVIQRMDDTKTGLVHTGSTPSSSTPFFNKEBSAILKFGAEL 977
 QY 769 FASEDDDEAGKSGKHYYDDAAIDKLLDRDLVBAE---VSDVDEENGFLKAFKVFYI 825
 DB 978 FKEPEGEQEP-----QEMDIDELKRAETRENEFGPLTVGDB-----LLSQKVFANFSNM 1028
 QY 826 DENEAAALEAORVAAESKSSAGNSDRASYWEEL-----LKDKFELHQAELNALGKR 877
 DB 1029 DEDD-IELEPER-----NSRWEEIIPVORRIIEEEROKELEIYMLPRM 1074
 QY 878 KRSRQLVSIIEEDDLALEDVSDGDESYEAESTDGEAAGQGVQTRPYRRKGRDNLBP 937
 DB 1075 RNCAQ-----ISFNGSEGRSRRSYSGSDSDSISERKPKKGRPRTP 1120
 QY 938 TPLMEGGRSFRVLGPNOSQRAIFVQTLRMRYGAGNFDWKEFVPR---LKQTFEINEYGV 994
 DB 1121 RE-----NIKGFSDAEIRRFIKSYKFG-GPLERLDVARDADAEVLDKSETDLRLIG 1170
 QY 995 ILFLKHAEEIDENS-----PTFSQVPEKGLRIBDLVRIALLILVOBK 1039
 DB 1171 ELVHNGCTALKNDSSGOERAGGRGLGVKGTFF-----RISGVQNAKLVLISHSEE 1221
 QY 1040 VKFEDHPGKVPFPPSRIER-----FPLRSGKIWKEEHDKTIRAVLKHGYGRWOA 1091
 DB 1222 LA-----PLHKSIPSDPEERKRYVIPCHTKAAHPDIDMGKEDDSNLLIGIYEGYSMEM 1276
 QY 1092 IVDDKELGIGIELCKELNPHISAAEQAGLQGGSGGSGNPGAGTQNPQSVITGNNN 1151
 DB 1277 IKMDPDLSTQKILPD-----DPDKKPAQAKQLQTRADYLIKLIN 1315
 QY 1152 ASADGAQVNSMFYFEDMORLVEFKKRVLLLEKAMNVEYAEYVYGLGSSSIPTPEEPA 1211
 DB 1316 KDLARKEAQLAGAGNSKRKTRTKKNKVLKAAK-LKEEIKSD-----SSQPSSEKSD- 1367
 QY 1212 EPKIADTVGVSFIEVDDMLDGLPKTDPTITSEEINGAAVDNNQARVEIAQHYNQMKCLID 1271
 DB 1368 -----EDDD-----NNKDEIVSVKHLHKKFAEK 1392
 QY 1272 ENARESVOAYVNNQPPSTKYVNESFALKSINGNINTILSITSDQSKSHEDDTKPOLNVE 1331
 DB 1393 ENEEK-----PEPDTGIKKE-----AEEKETKEKENKRDCLKREK 1427
 QY 1332 M-KDTAEETKPLRGVVDLNVVEGEENTAEASGVSDVKWEEAKEEEKPK 1379
 DB 1428 KEKEDKELKREKNKV-----KESTQKEKEVKEEKVNEIKSENK 1469

This Page Blank (uspto)